|  |  |  |
| --- | --- | --- |
|  |  |  |
| partitureditor2 | coma2 | exakt |
|  |  |  |
| **How to use an EXMARaLDA corpus** | | |
|  | | |

This document explains how to use EXMARaLDA corpora from [www.exmaralda.org](http://www.exmaralda.org/), [www.corpora.uni-hamburg.de](http://www.corpora.uni-hamburg.de/) and other sites. 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**

[A. Overview 2](#_Toc361821171)

[B. Online use in a Browser 2](#_Toc361821172)

[1. Corpus overview 2](#_Toc361821173)

[2. HTML visualizations 4](#_Toc361821174)

[C. Offline use 5](#_Toc361821175)

[D. Use as remote corpus in EXAKT 6](#_Toc361821176)

[1. Open a remote corpus 6](#_Toc361821177)

[2. Choose a corpus 7](#_Toc361821178)

[3. Login 7](#_Toc361821179)

[System version 1.5.1.]

# Overview

There are several ways to use an EXMARaLDA corpus. Which of these is preferable depends on what you want to do with the data (e.g. read it vs. query it), on where you want to use the data (e.g. on your office computer vs. a university course) and, last but not least, on the specifics of the data themselves. For example, the following usage scenarios are common:

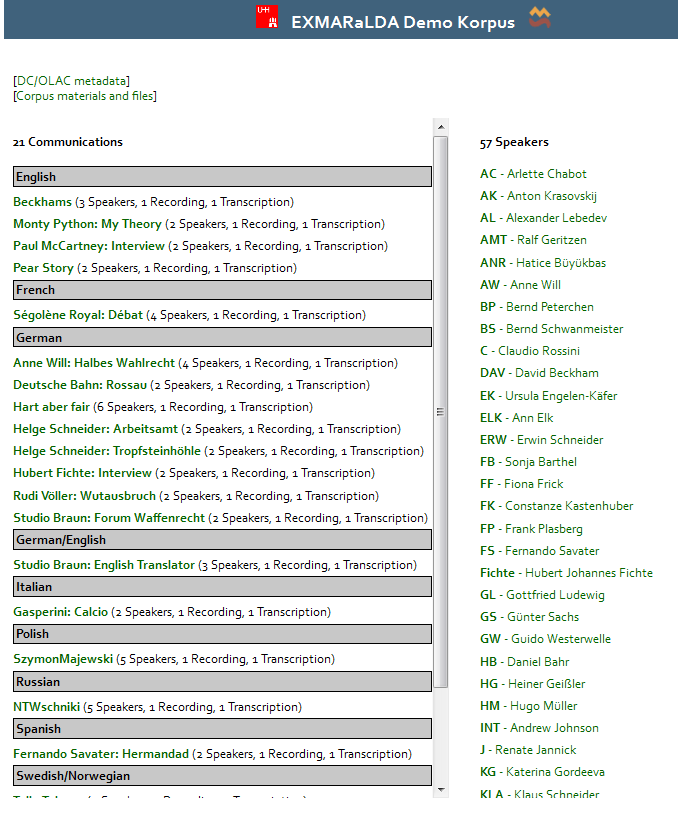
* You want to get an overview of the data in a corpus, listen to the recordings, read the transcriptions and inspect the metadata. For this scenario, it is usually recommended to browse the data online. This is explained in more detail in **section** **B**.
* You want to do a query on the data, e.g. you are looking for uses of a specific lexical item or all items which have been tagged or annotated in a certain way. For this scenario, you will want to use EXAKT. As explained in more detail in **section C**, EXAKT can access a corpus in three different ways – on a local file system (meaning you will have to download the data first), on a remote file system (meaning you can access the data directly on the server) or on a (remote) relational database.
* You want to make your own changes to the data, e.g. add a specific annotation to it or transfer it for use with some other tool. For this scenario, you will have to download the data. This is explained in more detail in **section D**.

# 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

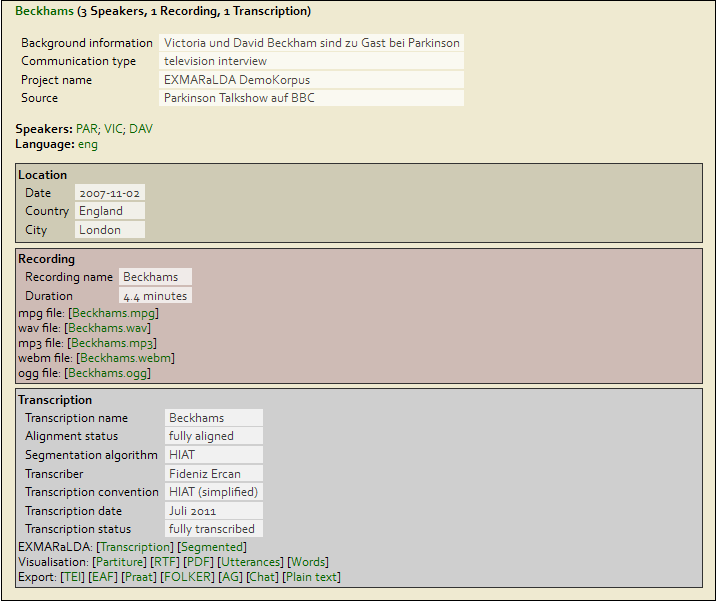
You usually start with a corpus overview. The corpus overview consists of a list of communications on the left and a list of speakers on the right side.



List of communications

List of speakers

By clicking on an item in one of these lists, information about the chosen item (the communication “Beckhams”) will be displayed:



2

1

3

3aaa

3baa

3caa

For communications, the top part gives you a list of metadata items (1). This is followed by a list of speakers participating in the communication (2). By clicking on any of the speakers, the corresponding information will be displayed in the speaker list.

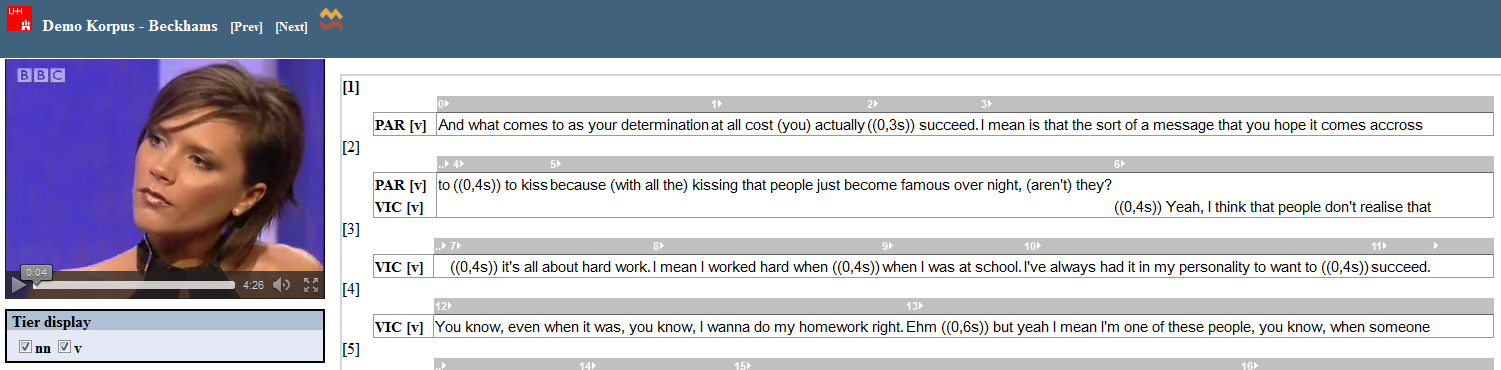
The lower part links to all the documents (recordings, transcriptions, visualisations and export formats) which belong to this communication (3). More specifically:

* The section EXMARaLDA (3a) links to an EXMARaLDA *basic transcription* (Transcription), which can be opened and edited in the EXMARaLDA Partitur-Editor, and an EXMARaLDA *segmented transcription* (Segmented), which is the file format used for querying in EXAKT.[[1]](#footnote-1)
* The section Visualisation (3b) links to *musical score* visualisations in different formats (Partiture, RTF, PDF), to an *utterance list* (Utterances) and a *word list* (Words) and – in older corpora – to a separate visualisation of the transcription head (HTML).
* The section Export (3c) links to several export formats. TEI is an XML file corresponding to the guidelines of the Text Encoding Initiative. 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. FOLKER is a file which can be opened with the FOLKER transcription editor of the IDS in Mannheim. AG is an annotation graph file that can be used for data exchange with various annotation tools. CHAT is the file format of the CLAN editor of CHILDES. Or, finally, you can also view the transcription in your browser in Plain text format.

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.

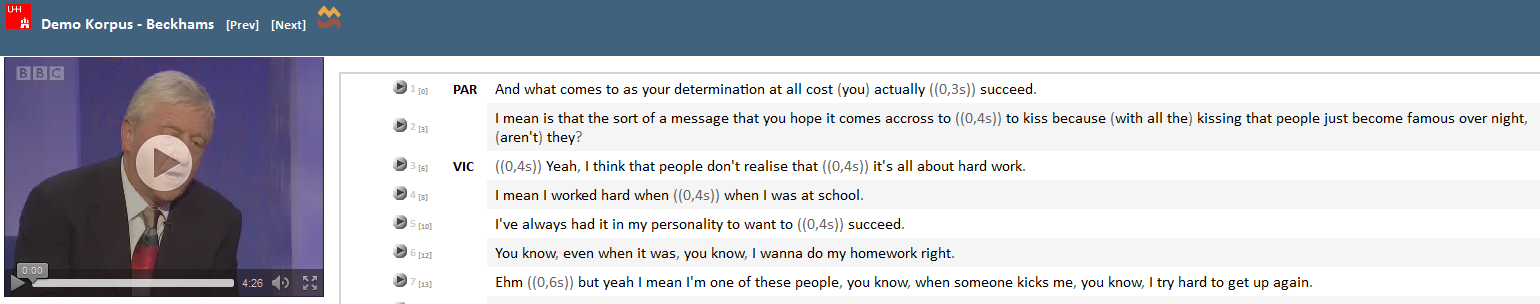
## HTML visualizations

If you display the HTML version (Partiture) 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 or, in newer corpora, to a HTML5 audio player:



Player

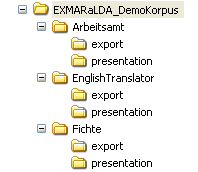
In the musical score, if you click on any one of the little arrows  in the top rows, the player will start playback at the corresponding time in the audio recording. Clicking on any number in the top rows of a musical score will take you to the corresponding place in an utterance list:



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

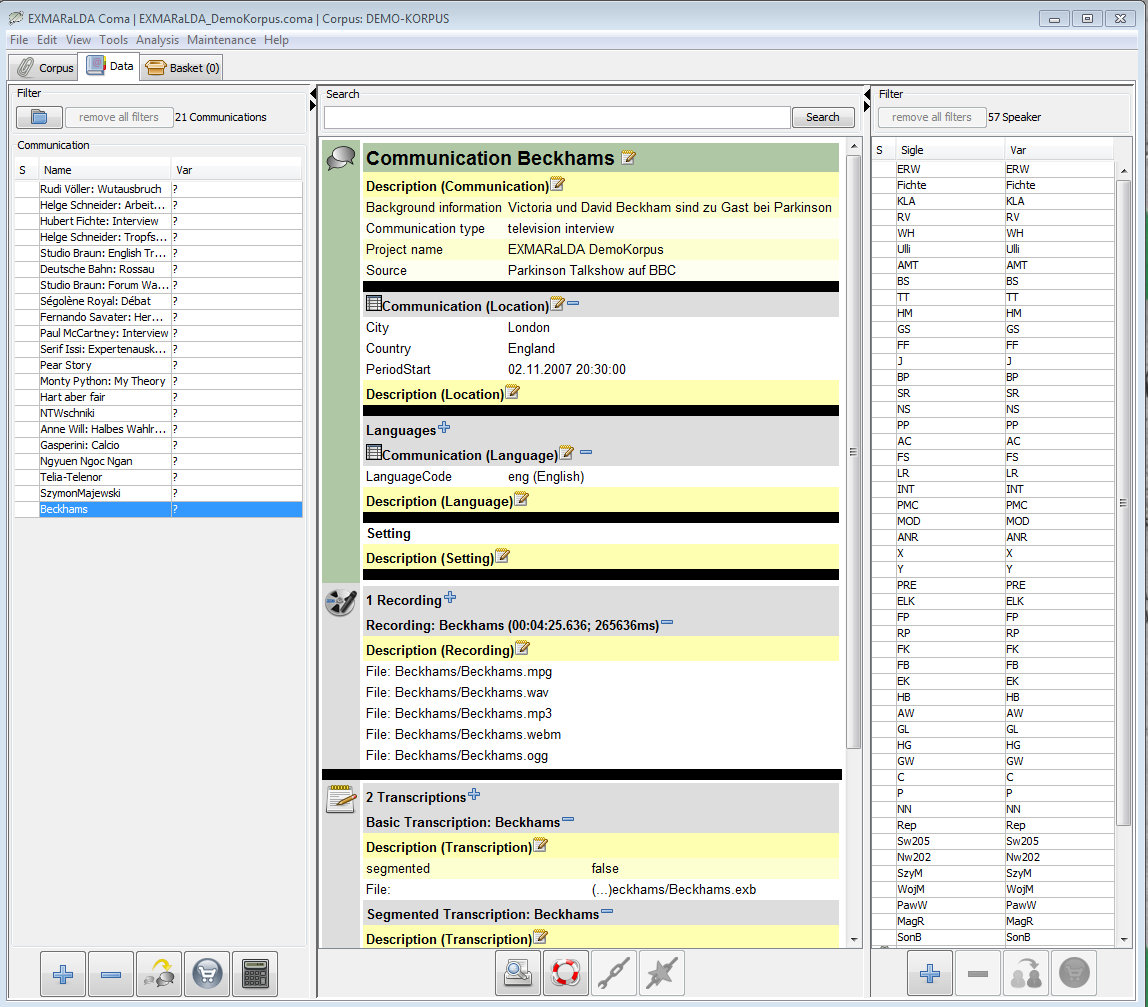
# Offline use

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 (from: [www.exmaralda.org](http://www.exmaralda.org/)) 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 a corpus file with the suffix ".coma" (for the demo corpus, this file is named EXMARaLDA\_DemoKorpus.coma[[2]](#footnote-2)).

The file can be opened with the Corpus Manager (Coma) to view, edit or query meta-data...

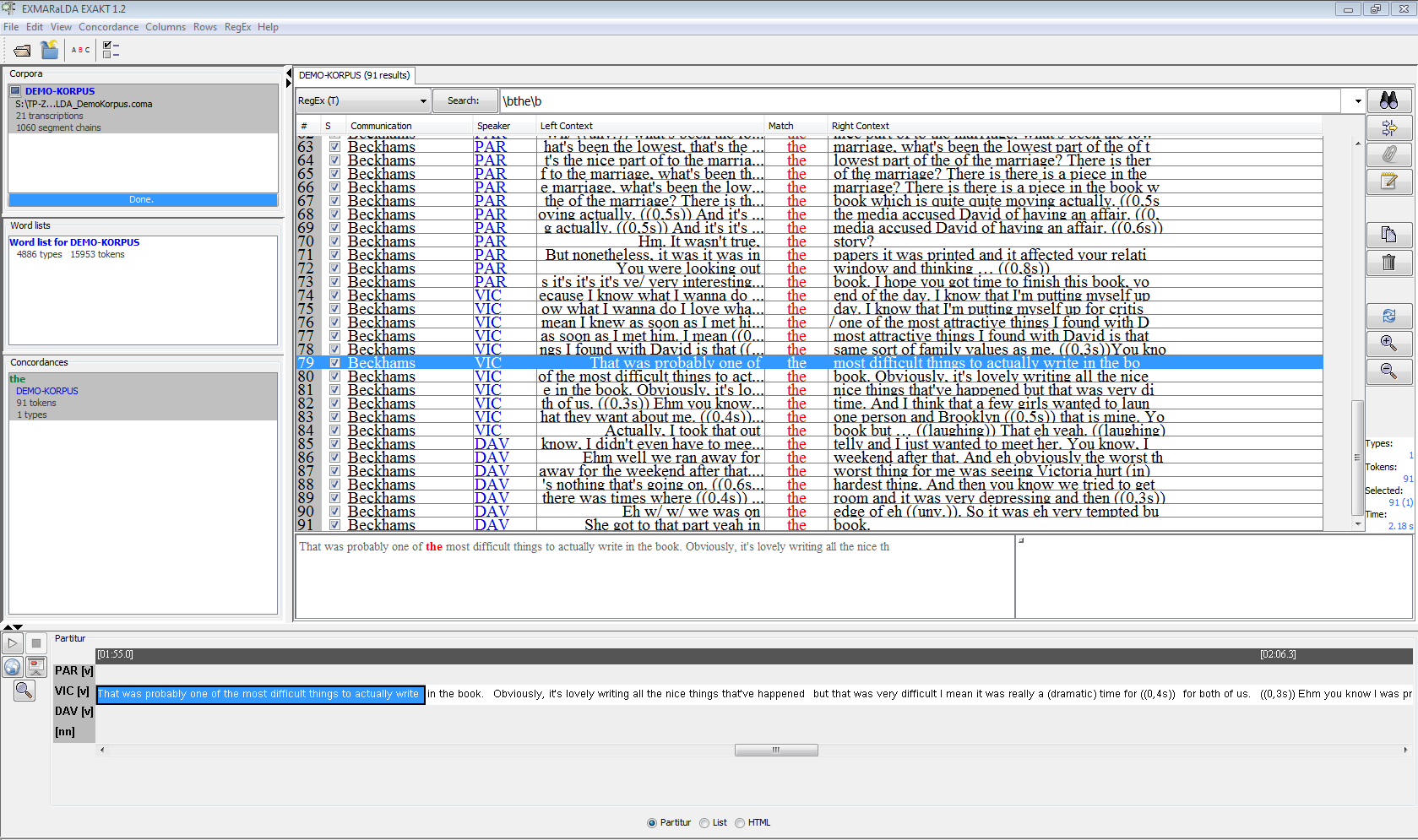


Corpus which is open at the moment

Click on one of the communications to view the corresponding metadata

Click on one of speakers to view the metadata

... or you can open it also with EXMARaLDA's query tool EXAKT to carry out corpus queries (from the menu File > Open corpus…).



Here you can see which corpora are open

Wordlists

Concordance(s)

Query

Partiture view in EXAKT

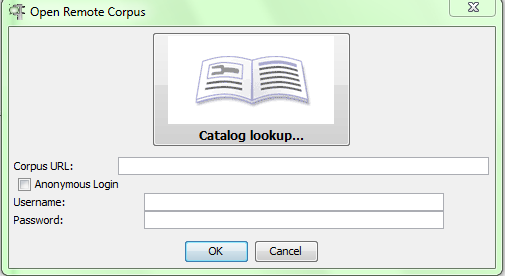
Of course, you can also view and edit individual transcription files (you'll find those inside the subfolders, they have the suffix ".exb" or – in older corpora ".xml") with the EXMARaLDA Partitur-Editor.

For further information on how to query an EXMARaLDA corpus, please consult the Coma and EXAKT documentation.

# Use as remote corpus in EXAKT

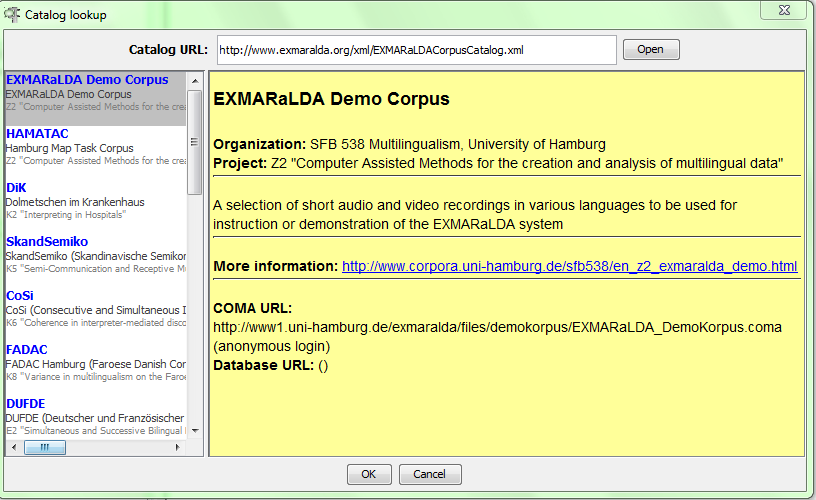
EXAKT offers the possibility to do queries on a remote corpus, i.e. on a corpus that is not located on your own computer, but on a remote server. Note that you normally need a broadband connection for this feature to work satisfactorily (other connections are too slow).

## Open a remote corpus

Start EXAKT and choose File > Open remote corpus...

In the following dialog, you have to click on the book symbol to browse through available corpora.

## Choose a corpus

A window with the corpora from the SFB ‘Multilingualism’ will appear automatically. Choose an entry (left menu) and click OK – the correct URL will be entered for you:

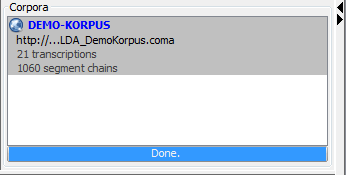
Click on the corpus to view the information in the right menu

Click here to confirm your choice

## Login

Afer you have chosen the corpus, a dialog with login request (just like in step 1 “Open a remote corpus”) will appear.

1. If the corpus is password protected, you have to enter your username and password in the fields Username and Password, respectively and then click OK. Please note: Check the website [www.corpora.uni-hamburg.de](http://www.corpora.uni-hamburg.de/)[[3]](#footnote-3) (Section “Ressourcen und Projekte” > “Korpora”) for access permission and/or contact information.
2. 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 (compare also: screenshot in **section C**):



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.

1. For more information about the two formats, consult “Quickstart\_Segmentation” and “How to use segmentation”. [↑](#footnote-ref-1)
2. Older corpora use the suffix ".xml" instead of ".coma" [↑](#footnote-ref-2)
3. https://corpora.uni-hamburg.de/drupal/en/islandora/object/collection%3Asfb632. [↑](#footnote-ref-3)