|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
| **How to:**  **Verwendung eines EXMARaLDA-Korpus** | | |
|  | | |

Dieses Dokument erläutert, wie Sie EXMARaLDA-Korpora von [www.exmaralda.org](http://www.exmaralda.org/),[www.corpora.uni-hamburg.de](http://www.corpora.uni-hamburg.de/) oder anderen Seiten verwenden können. Die Beispiele basieren auf dem EXMARaLDA-Demokorpus. Andere Korpora funktionieren auf eine ähnliche Weise, können sich aber bezüglich der Verfügbarkeit von Audiodaten oder Typen von Exportformaten unterscheiden.

**Inhalte**

[A. Überblick 2](#_Toc361819788)

[B. Online-Nutzung im Browser 2](#_Toc361819789)

[1. Korpusübersicht 2](#_Toc361819790)

[2. HTML Visualisierung 4](#_Toc361819791)

[C. Offline-Nutzung 5](#_Toc361819792)

[D. Nutzung als Remote-Korpus in EXAKT 7](#_Toc361819793)

[1. Einen Remote-Korpus öffnen 7](#_Toc361819794)

[2. Korpus wählen 7](#_Toc361819795)

[3. Login 8](#_Toc361819796)

[Systemversion 1.5.1.]

# A. Überblick

Es gibt verschiedene Möglichkeiten, ein EXMARaLDA-Korpus zu nutzen. Welche vorzuziehen ist, hängt davon ab, was man mit den Daten machen möchte (z.B. lesen oder gezielt durchsuchen) und wo die Daten genutzt werden sollen (z.B. auf dem Computer am Arbeitspatz oder in einem Seminar). Nicht zuletzt sind die Eigenschaften der Daten selbst für die Art der Verwendung von Bedeutung. Folgende Nutzungsszenarien sind häufig:

* Man möchte sich einen Überblick über die Daten in einem Korpus verschaffen, die Aufnahmen anhören, Transkripte lesen und die Metadaten anzeigen. Für dieses Szenario ist es empfehlenswert, die Daten online zu durchsuchen. Hierauf wird im **Abschnitt B** ausführlicher eingegangen.
* Eine Datensuchabfrage soll durchgeführt werden: z.B. man will sich die verschiedenen Verwendungsformen einer bestimmten lexikalischen Einheit oder aller Einheiten ansehen, die annotiert oder mit speziellen „Tags“ versehen worden sind. Hierfür ist es am besten EXAKT zu verwenden. Wie im **Abschnitt C** genauer erläutert wird, kann EXAKT auf drei verschiedene Arten auf ein Korpus zugreifen – auf einem lokalen Dateisystem (Download erforderlich), auf einem entfernten Dateisystem (hierbei können die Daten direkt auf dem Server abgerufen werden) oder auf einer (entfernten) relationalen Datenbank.
* Eigene Änderungen an den Daten sollen vorgenommen werden, z.B. eine bestimmte Annotation hinzufügen oder die Daten für die Benutzung mit einem anderen Tool übertragen. Hierfür müssen die Daten heruntergeladen werden. Darauf wird in **Abschnitt D** ausführlicher eingegangen.

# B. Online-Nutzung im Browser

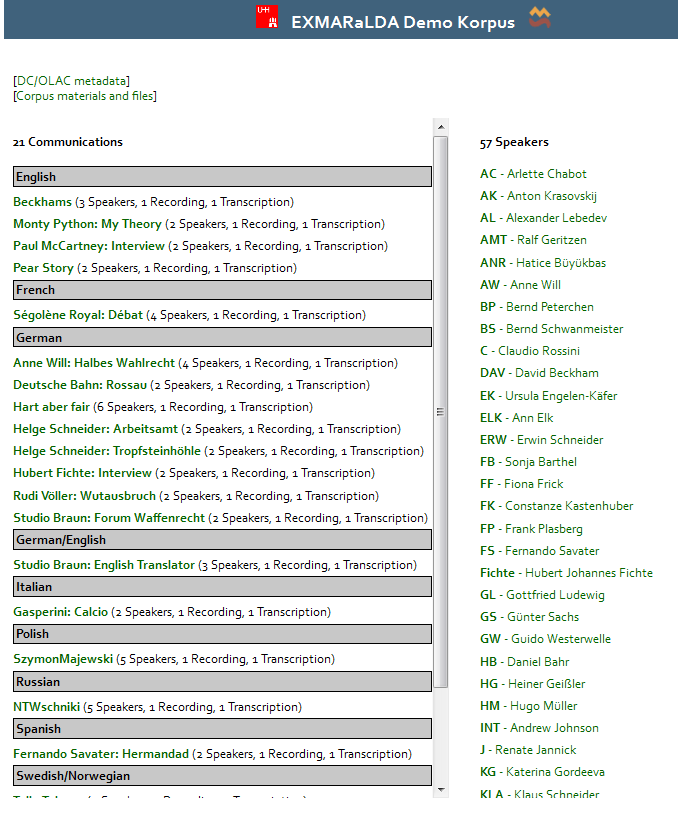
Das Korpus kann online benutzt werden (d.h. im Internet-Browser, ohne die Daten herunterzuladen) um die Metadaten und Transkriptionen anzuzeigen und ggf. Aufnahmen abzuspielen.

## Korpusübersicht

Normalerweise beginnt man mit der Korpusübersicht. Diese besteht aus einer Liste von Kommunikationen (links) und einer Liste von Sprecher (rechts).

Kommunikationsliste

Sprecherliste



Durch einen Klick auf einen Eintrag in einer der beiden Listen werden Informationen zu dem Eintrag angezeigt (hier die Kommunikation „Rudi Völler: Wutausbruch“):

1

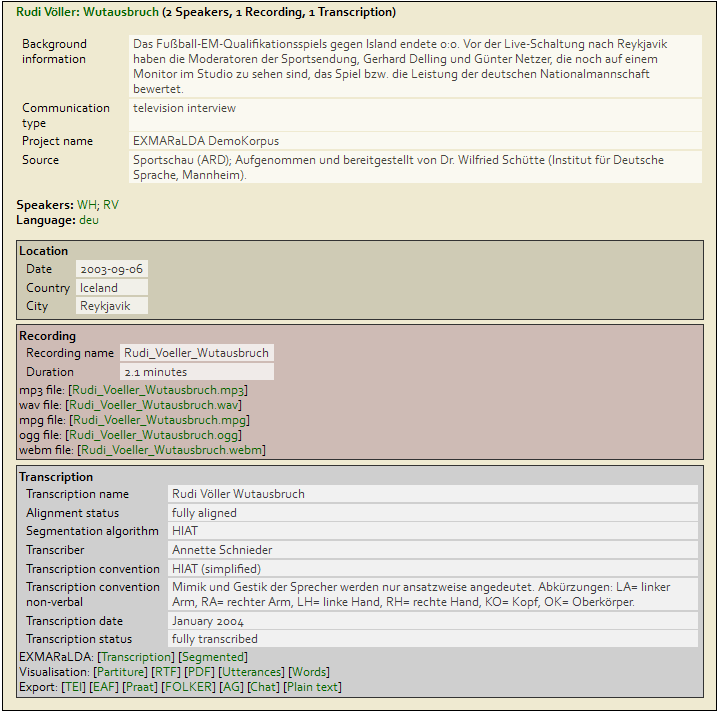
2

3

3aaa

3baa

3caa



Bei einer Kommunikation finden Sie im oberen Teil eine Liste mit Metadateneinträgen (1). Dieser folgt eine Liste von Sprechern, die an der Kommunikation beteiligt sind (2). Durch einen Klick auf einen Sprecher, wird die entsprechende Information zum Sprecher in der Sprecherliste angezeigt.

Der untere Teil führt alle Dokumente (Aufnahmen, Transkripte, Visualisierungen und Exportformate) auf, die mit der jeweiligen Kommunikation verknüpft sind (3). Genauer:

* Im Abschnitt EXMARaLDA (3a) sind jeweils Links zu einer EXMARaLDA-*Basistranskription* (Transcription), die mithilfe des Partitur-Editors geöffnet und bearbeitet werden kann, sowie zu einer *segmentierten Transkription* (Segmented), das Dateiformat, welches für Suchabfragen in EXAKT benötigt wird.[[1]](#footnote-1)
* Der Abschnitt Visualisation (3b) verfügt über eine Verlinkung zu Visualisierungen in Form einer *Partitur* in verschiedenen Formaten (Partiture, RTF, PDF), zu einer *Äußerungsliste* (Utterances), zu einer *Wortliste*(Words), sowie – bei älteren Korpora – zu einer separaten Visualisierung des Transkriptionskopfes (HTML).
* Über den Abschnitt Export gelangt man zu verschiedenen Export-Formaten. TEI ist eine XML-Datei, entsprechend den Richtlinien der Text Encoding Initiative. EAF ist eine ELAN-Annotationsdatei, die mithilfe von ELAN-Werkzeugen der MPI in Nijmegen geöffnet und bearbeitet werden kann. Praat ist eine TextGrid-Datei, das mithilfe der Praat-Software geöffnet und bearbeitet werden kann. FOLKER ist eine Datei, die mit dem FOLKER-Transkriptionseditor des IDS in Mannheim verwendet werden kann. AG ist eine „Annotation Graph“-Datei, die für den Datenaustausch mit verschiedenen Annotationswerkzeugen verwendet werden kann. CHAT ist das Dateiformat des CLAN-Editors von CHILDES. Schließlich kann die Transkription auch als Plain text direkt im Browser geöffnet werden.

Klicken Sie auf eine dieser Dateien, um sie in Ihrem Browser anzeigen zu lassen. Um die Dateien herunterzuladen, müssen Sie möglicherweise über die rechte Maustaste die Option „Speichern unter...” aus dem Kontextmenü auswählen.

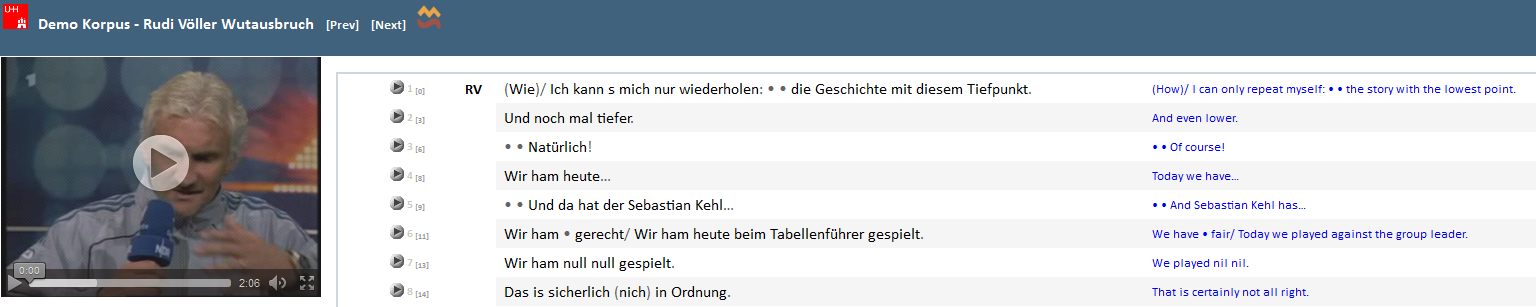
## HTML-Visualisierung

Lässt man sich die HTML-Version der Partitur (Partiture) anzeigen (und sind die Audio-Aufnahmen im Korpus verfügbar), bekommt man eine Transkription, die entweder mit einem Flash Audio Player oder, in neueren Korpora, mit einem HTML 5 Audio Player verknüpft ist:

Player



Durch Klick auf einen der kleinen Pfeile in den oberen Reihen der Partitur, startet der Player die Wiedergabe an der entsprechenden Stelle in der alignierten Audioaufnahme. Der Klick auf eine Zahl in den oberen Reihen der Partitur führt an die entsprechende Stelle in der Äußerungsliste:

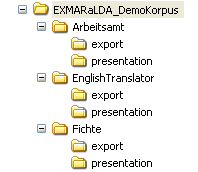


Auch hier wird durch einen Klick auf den Pfeil neben der Äußerung der Audio Player gestartet. Mit einem Klick auf eine Zahl in eckigen Klammern gelangt man zurück zur entsprechenden Stelle der Partitur.



# C. Offline-Nutzung

Für die Offline-Nutzung kann das gesamte Korpus heruntergeladen werden. Dies ist besonders dann nützlich, wenn man die Daten selbst bearbeiten oder Korpusabfragen durchführen will. Um das Korpus herunterzuladen, reicht ein Klick auf den Link zum ZIP-Archiv (auf [www.exmaralda.org](http://www.exmaralda.org/)). Nach Entpacken der Datei auf der Festplatte sollte folgende Verzeichnisstruktur vorhanden sein:



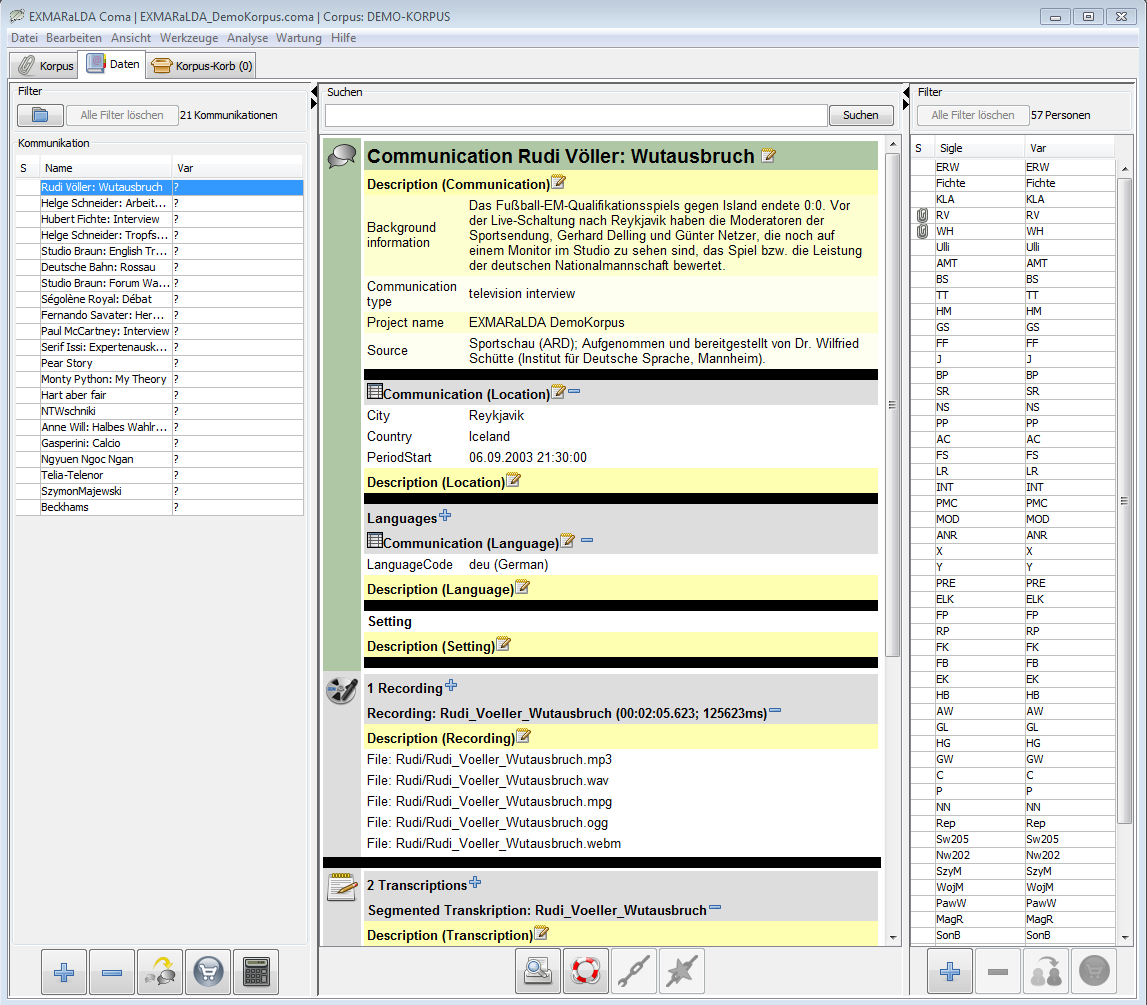
Auf der obersten Verzeichnisebene müsste nun eine Korpusdatei mit dem Suffix ".coma" zu finden sein (für das Demokorpus lautet der Dateiname EXMARaLDA\_DemoKorpus.coma[[2]](#footnote-2)).

Diese Datei kann nun mit dem EXMARaLDA Corpus Manager (Coma) geöffnet werden, um Metadaten anzuzeigen, zu bearbeiten oder zu durchsuchen…

Durch Klick auf einen der Sprecher werden die entsprechenden Metadaten eingeblendet

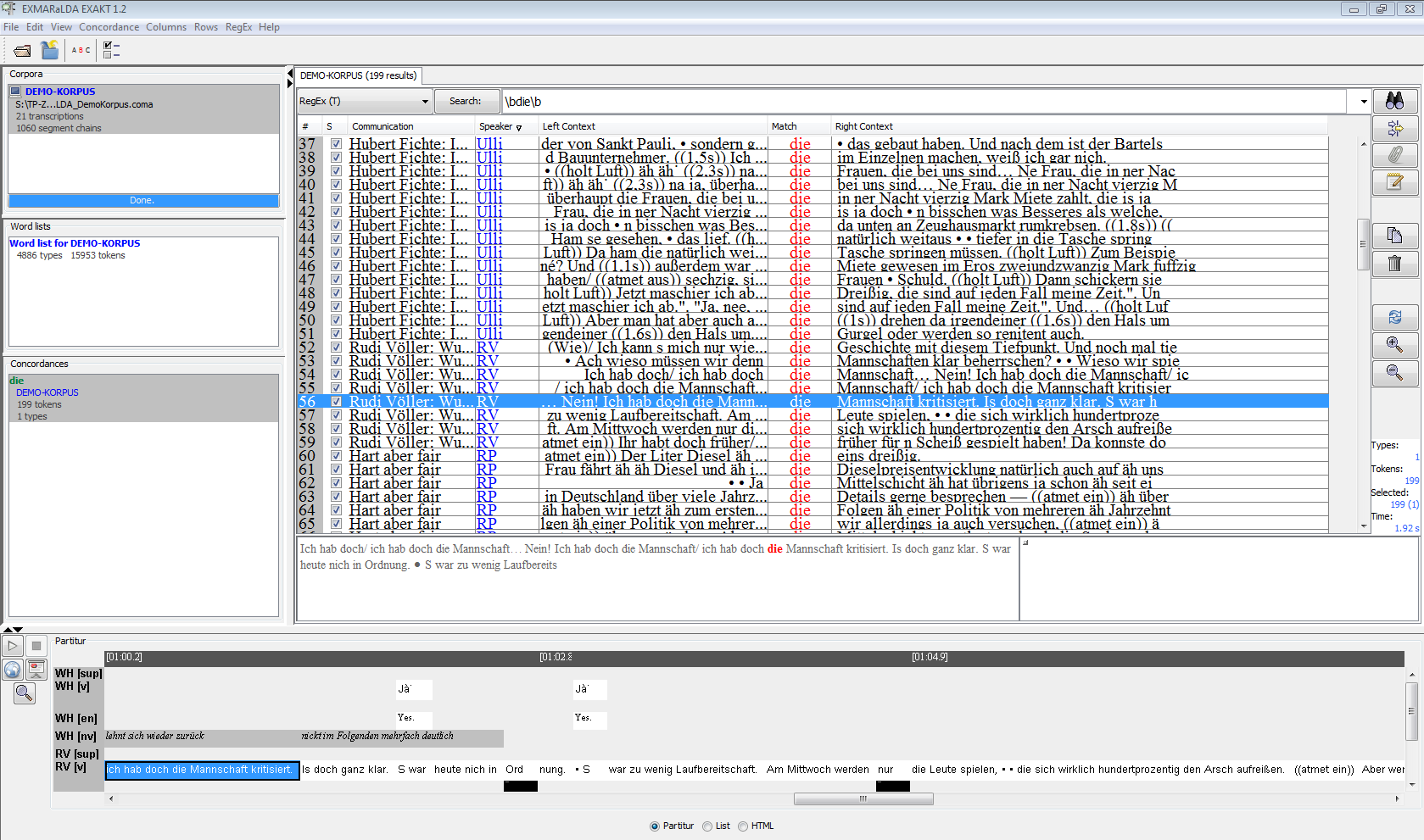
Durch Klick auf eine der Kommunikationen werden die dazugehörigen Metadaten eingeblendet

Das momentan geöffnete Korpus



... oder mit dem EXMARaLDA-Suchwerkzeug EXAKT, um Korpusabfragen auszuführen (über das Menu File > Open corpus…):

Suchabfrage (Query)



Liste der geöffneten Korpora

Wortlisten

Konkordanz(en)

Partituransicht in EXAKT

Die einzelnen Transkriptionsdateien (diese sind mit dem Suffix ".exb" oder – in älteren Korpora – ".xml" in den Unterordnern zu finden) können mit dem EXMARaLDA Partitur-Editor angesehen und bearbeitet werden.



Weitere Informationen zu Korpus-Suchabfragen sind in der Dokumentation zu Coma und EXAKT zu finden.

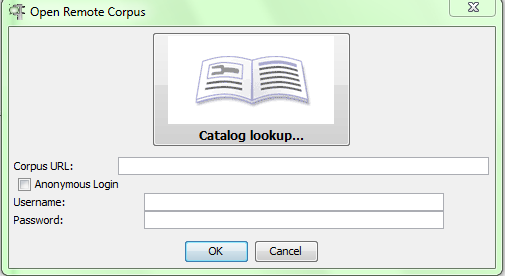
# D. Nutzung als Remote Corpus in EXAKT

EXAKT bietet die Möglichkeit, Suchabfragen an einem Remote Corpus durchzuführen, also an einem Korpus, das sich nicht auf dem eigenen Computer, sondern auf einem entfernten Server befindet. Dabei ist zu beachten, dass, um mit dieser Funktion zufriedenstellend arbeiten zu können, ein Breitbandzugang benötigt wird, da andere Anschlüsse zu langsam sind.

Partiture view in EXAKT

## Ein Remote Corpus öffnen

EXAKT starten und File > Open remote corpus... wählen:



Partiture view in EXAKT

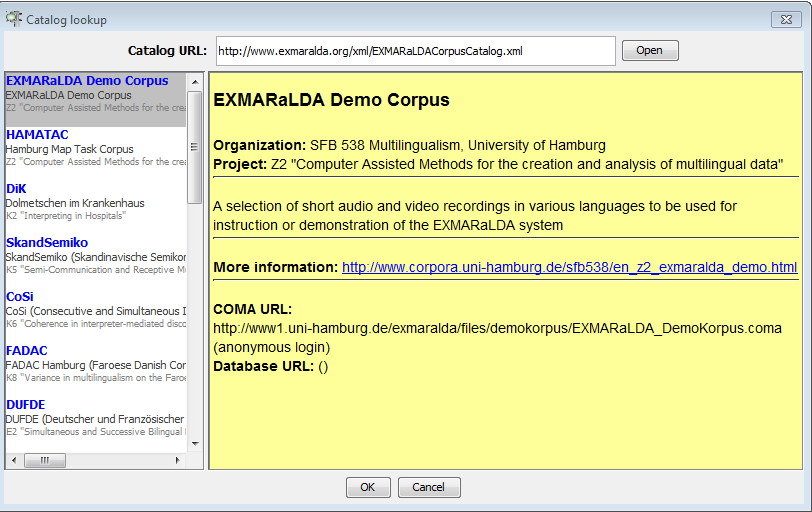
Im folgenden Dialogfenster auf das Buchsymbol klicken, um nach vorhandenen Korpora zu suchen

## Korpus wählen

Ein Fenster mit den Korpora vom SFB ‘Mehrsprachigkeit’ erscheint automatisch. Nun kann das Korpus ausgewählt werden (linke Menüspalte). Durch einen Klick auf OK wird die entsprechende URL automatisch übernommen:

Durch einen Klick wird die Auswahl übernommen

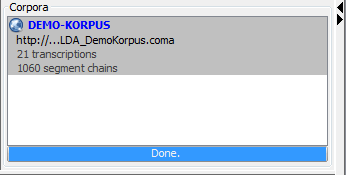
Durch Klick auf einen Korpus werden die Informationen in der rechten Menüspalte angezeigt



## Login

Als Nächstes wird ein Fenster mit einer Login-Aufforderung erscheinen (genau wie im Schritt 1 „Ein Remote Corpus öffnen“ – im Screenshot).

1. Wenn das Korpus passwortgeschützt ist, müssen Benutzernamen und das Passwort in die entsprechenden Felder UsernameundPassword eingeben werden. Dann auf OK klicken. Detaillierte Inforamtionen zu Zugangsrechten und/oder Kontakdaten sind für jedes Korpus auf der Webseite [www.corpora.uni-hamburg.de](http://www.corpora.uni-hamburg.de/)[[3]](#footnote-3) (Sektion „Ressourcen und Projekte“ > „Korpora“) zu finden.
2. Wenn das Korpus nicht passwortgeschützt ist (wie im Fall des EXMARaLDA-Demokorpus), die Option Anonymous loginwählen, um die Benutzerauthentifizierung zu deaktivieren. Anschließend auf OK klicken, um das Korpus zu öffnen, welches dann auch in der Liste der geöffneten Korpora links oben im EXAKT Fenster angezeigt wird (vgl. dazu auch den Screenshot in **Sektion C**):



Man kann mit diesem Korpus dann arbeiten, als wäre es auf dem lokalen Computer. Zu beachten ist jedoch, dass das Abspielen von Audio-/Videodateien z.Zt. nur von Windows unterstützt wird. Auf Macintosh und Linux können lediglich die Mediadaten der lokalen Korpora abgespielt werden.



1. Für weitere Informationen über die beiden Formate konsultieren Sie „Quickstart Segmentation“ und „How to use segmentation“. [↑](#footnote-ref-1)
2. Ältere Korpora nutzen die Dateierweiterung „.xml“ anstelle von „.coma“ [↑](#footnote-ref-2)
3. Diese Webseite ist nur in Deutsch verfügbar. [↑](#footnote-ref-3)