

EXMARaLDA Corpus Manager

Manual

Version 1.6

Last update: October, 2016

Main contributor: Kai Wörner

Further contributors: Anne Ferger, Carolin Frontzek,
Karolina Kaminska, Heidemarie Sambale

Inhaltsverzeichnis

1	Über Coma	4
2	Die Coma-Datenstruktur	5
2.1	Korpus	5
2.2	Kommunikationen	5
2.3	Sprecher	5
2.4	Aufnahmen	6
2.5	Transkriptionen	6
2.6	Verknüpfte Dateien	6
2.7	Weitere Datentypen	6
2.7.1	Locations	7
2.7.2	Description	7
2.7.3	Anmerkung zu selbstgewählten Metadatenschlüsseln	7
3	Die Programmoberfläche	8
3.1	Reiter	8
3.2	Menüs	8
3.2.1	Datei-Menü	8
3.2.2	Bearbeiten-Menü	10
3.2.3	Ansicht-Menü	10
3.2.4	Werkzeuge-Menü	10
3.2.5	Analyse-Menü	10
3.2.6	Wartung-Menü	11
3.2.7	Hilfe-Menü	12
4	Korpusverwaltung.....	13
4.1	Korpus aus Transkriptionen erstellen	14
4.1.1	Verwenden des Assistenten	14
4.2	Importieren und Exportieren von Metadaten	19
4.2.1	Basistranskriptionen Importieren	19
4.2.2	Sprecher importieren	19
5	Verwaltung von Metadaten	21
5.1	Aufbau des Daten-Reiters	21
5.2	Verwalten von Kommunikationen	22
5.3	Metadaten-Anzeige	23
5.4	Eingabedialoge	25
5.4.1	Location	25
5.4.2	Language	26
5.5	Templates	27
5.5.1	Templates speichern	28
5.6	Verwalten von Personen und Sprechern	28
5.6.1	Importieren von Sprechern	28

5.7	Verknüpfen von Kommunikationen und Sprechern.....	29
5.8	Verwalten von Transkriptionen.....	29
5.9	Verwalten von Aufnahmen.....	30
5.10	Verwaltung von zusätzlichen Dateien	31
6	Analysieren von Daten	32
6.1	Filter	32
6.1.1	Filter anlegen.....	32
6.1.2	Filter modifizieren.....	33
6.1.3	Vorgefertigte Filter.....	33
6.1.4	Filter löschen	33
6.1.5	Filter bei Sprechern	33
6.2	Suchen	33
6.3	In Transkriptionen suchen	34
6.4	Drucken der Metadatenansicht	34
6.5	Korpusstatistik erstellen	35
6.6	Wortliste erstellen.....	36
6.7	Wortliste für ausgewählte Sprecher erstellen	37
6.8	Ausgabe als html-Datei	37
6.9	Korpus-Korb.....	37
6.9.1	Verwalten des Korbes	37
6.10	TreeTagger	38
7	Wartung	40
7.1	Überprüfung verknüpfter Transkriptionen	40
7.1.1	Transkriptionen segmentieren.....	40
7.1.2	Auf Strukturfehler überprüfen.....	40
7.1.3	Fehlerliste	41
7.1.4	Auf Segmentierungsfehler überprüfen	42
7.1.5	Sprecherzuordnung prüfen	42
7.1.6	Suchen und Ersetzen im Korpus	42
7.2	Pflege der Metadaten	43
7.2.1	Segmentzahlen aktualisieren	43
7.2.2	Aufnahmen aus Transkriptionen aktualisieren.....	43
7.2.3	Description-Schlüssel angleichen	43
8	Einstellungen.....	45
9	Anmerkungen und Haftung	47
9.1	Haftungsbegrenzung.....	47
9.2	Preview-Versionen	47
9.3	Dieses Dokument.....	47

1 Über Coma

Der **EXMARaLDA Corpus-Manager** (Coma) ist ein Softwarewerkzeug, mit dem EXMARaLDA-Transkriptionen und die dazu gehörigen Aufnahmen mit relevanten Metadaten versehen und zu Korpora zusammengefasst werden. Unter anderem beinhaltet der Corpus-Manager folgende Funktionen:

- Eingabe und Verwaltung von Informationen zur Kommunikation, zu den beteiligten Personen, zu den Aufnahmen und zu den Transkriptionen;
- Organisation dieser Informationen und Verknüpfung mit den Aufnahme- und Transkriptionsdateien;
- Filtern von Metadaten und Zusammenstellung von Teil-/Untersuchungskorpora anhand der Metadaten. Diese Teilkorpora dienen dann als Ausgangspunkt für das Suchwerkzeug EXAKT;
- Hilfe bei der Erstellung neuer Transkripte mithilfe des Partitureditors und der Funktion „**Neu aus Assistenten...**“ (siehe http://www.exmaralda.org/pdf/PartiturEditor_Handbuch.pdf).

Bitte beachten Sie: Das Programm befindet sich in einer fortwährenden Weiterentwicklung („Work in Progress“). Informationen über aktuelle Änderungen finden Sie unter <http://www.exmaralda.org/files/changes.html>.

2 Die Coma-Datenstruktur

Die Struktur, in der Metadaten in Coma verwaltet werden, kennt einige vorgegebene „Datencontainer“:

- Korpora,
- Kommunikationen,
- Sprecher,
- Transkriptionen,
- Aufnahmen,
- und verknüpfte Dateien.

Unter diesen „Containern“ existieren verschiedene Datentypen, die in mehreren dieser Container Verwendung finden können. Da es wichtig ist, das Zusammenspiel dieser Container zu verstehen, sollen sie hier kurz beschrieben werden:

2.1 Korpus

Ein **Korpus** bildet den Container für alle anderen Metadatentypen. Ein Korpus enthält **Metadaten** (zu Kommunikationen und Sprechern) sowie Verweise auf alle zum Korpus gehörenden Dateien.

2.2 Kommunikationen

Mit **Kommunikationen** werden Gesprächsereignisse verwaltet. An Kommunikationen nehmen typischerweise **Sprecher** teil und es kann **Aufnahmen** und **Transkriptionen** zu den Kommunikationen geben. Das heißt im Datenmodell können Aufnahmen, Transkriptionen, Dateien und Sprecher diesen Kommunikationen zugeordnet sein. Mit den Kommunikationen werden alle Konstellationsdaten zu dem Gesprächsereignis gespeichert, also Dinge wie Ort, Zeit, besondere Umstände, beteiligte Sprachen etc.

2.3 Sprecher

Sprecher sind Personen, die sich an der Kommunikation beteiligen. Es muss sich dabei nicht um natürliche Personen handeln (es kann beispielsweise auch ein automatisiertes Dialogsystem als Sprecher geführt werden) und die Sprecher müssen in der Kommunikation auch nicht unbedingt sprechen; wenn sie für die Kommunikation von Belang sind, werden sie hier erfasst. Die Daten, die zu den Sprechern erfasst werden, sind von der Kommunikation zunächst unabhängig. Es sollten keine kommunikationsbezogenen Daten mit den Sprechern erfasst werden, damit Sprecher mehreren Kommunikationen zugeordnet werden können. Erfasst werden dagegen Daten wie: Geburtsort und -Datum, biographische Angaben, gesprochene Sprachen und deren Status etc.

2.4 Aufnahmen

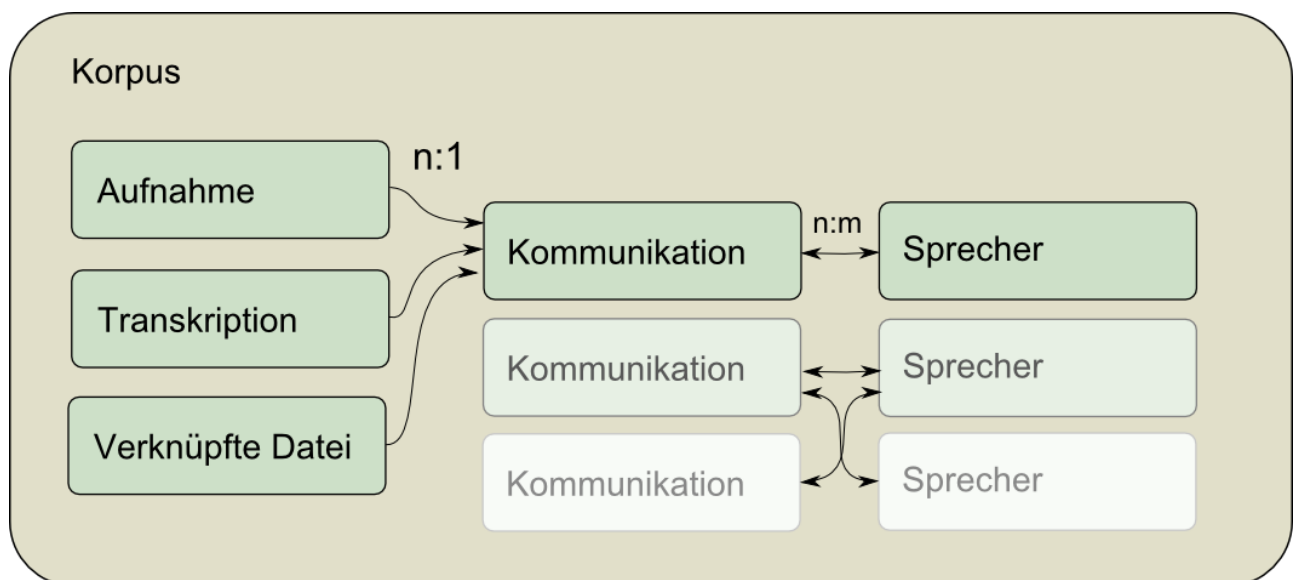
Aufnahmen beziehen sich auf Audio- oder Videoaufnahmen, die zu Kommunikationen angefertigt wurden. Das Coma-Datenmodell unterscheidet: Aufnahmen und tatsächliche Mediendateien. Einer Aufnahme können mehrere tatsächliche Mediendateien zugeordnet werden. Aufnahmen lassen sich nur anlegen, wenn es bereits eine Kommunikation gibt, zu der sie sich zuordnen lassen.

2.5 Transkriptionen

Im Datentyp **Transkriptionen** werden Verknüpfungen zu EXMARaLDA-Transkriptionen gespeichert. Coma verwaltet dabei sowohl Basis- (Dateiendung „.exb“) als auch segmentierte Transkriptionen (Dateiendung „.exs“). Zur Anzeige von Basistranskriptionen und zu ihrer Behandlung lassen sich in den Einstellungen von Coma Vorgaben machen (**Bearbeiten > Einstellungen**, s. auch Abschnitt *Einstellungen*). Transkriptionen sind, genau wie Aufnahmen, mit Kommunikationen verknüpft. Sie lassen sich also ohne Kommunikation nicht verwalten. Auch mit den Aufnahmen sind sie nur über die Kommunikation verbunden.

2.6 Verknüpfte Dateien

Zu Kommunikationen lassen sich auch **Dateien** zuordnen, die keine Aufnahmen oder Transkriptionen sind. Dies können beispielsweise: Aufnahmeprotokolle, Pläne, gescannte Metadaten-Erfassungsbögen oder ähnliches sein.



2.7 Weitere Datentypen

Zur Erfassung der eigentlichen Metadaten existieren noch weitere Datentypen, z.B. zur einheitlichen Erfassung von Sprachen und Orten. Zwei dieser Datentypen sind dabei von besonderer Bedeutung:

2.7.1 Locations

Eine **Location** bezeichnet in Coma nicht einfach nur einen bestimmten Ort, sondern *einen bestimmten Ort zu einer bestimmten Zeit*, da biographische und andere relevante Ereignisse häufig einen bestimmten Ort und eine bestimmte Zeit haben.

Geburt (Location)	
City	Hanau
Country	Deutschland
PeriodStart	13.04.1960 00:00:00
Description (Location)	

In dem oben angegeben Fall wird mit einer „**(Location)**“ beispielsweise Geburtsdatum und -Ort eines Sprechers erfasst. Es müssen in Locations nicht immer Ort UND Zeit vermerkt werden, es kann auch nur ein Zeitpunkt oder nur ein Ort angegeben werden. Es ist aber wichtig zu beachten, dass man auch dann eine Location verwenden muss, wenn man nur einen Zeitpunkt vermerken will.

2.7.2 Description

Da es extrem schwierig ist, standardisierte Metadaten für alle erdenklichen Fälle vorzusehen, wird in Coma der überwiegende Teil der Metadaten mit freien Schlüssel-Wert-Paaren erfasst. Diese Paare werden unter dem Datentyp **Descriptions** zusammengefasst. Sie existieren für die meisten der in Coma existierenden Datentypen: So lassen sich Descriptions für Korpora, Kommunikationen, Sprecher, Locations, Aufnahmen etc. anlegen.

Speaker: KLA (KLA, Sex: male)	
Description (Speaker)	
Ausbildung: beruflich	Friseur
Ausbildung: schulisch	Realschule
Beruf der Mutter	Hausfrau
Beruf des Vaters	Schlossermeister
Beruf	Arbeitslos
Hobbies	Speläologie
Name	Schneider
Vorname	Klaus

In diesem Beispiel sind in einer „**Description (Speaker)**“ Metadaten zu einem Sprecher erfasst.

2.7.3 Anmerkung zu selbstgewählten Metadatenschlüsseln

Dass sich die Felder innerhalb von Descriptions frei benennen lassen, macht sie nicht beliebig: Jeder Anwender sollte sich für sein Korpus ein festes Inventar an Description-Schlüsseln ausdenken und dokumentieren, um später mit den Metadaten auch gewinnbringend arbeiten zu können. Die Dokumentation dieser Metadatenschlüssel und ihrer möglichen Werte kann beispielsweise in Form eines Handbuchs geschehen.

Metadatenschlüssel werden in Coma automatisch alphabetisch sortiert. Um zusammengehörige Schlüssel nacheinander zu gruppieren, bietet es sich an ein passendes Präfix vor den Metadatenschlüssel zu setzen. Im obigen Beispiel wurde dies bei „**Ausbildung: schulisch**“ und „**Ausbildung: beruflich**“ angewendet.

3 Die Programmoberfläche

3.1 Reiter



Die meisten Funktionen von Coma lassen sich direkt im Programmfenster vornehmen. Das Fenster wird von drei Reitern gegliedert, über die sich die Funktionen für einzelne Bereiche des Programms erreichen lassen.

- **Korpus**-Reiter führt Sie zur Verwaltung der Korpora. Hier lassen sich die Metadaten, die das Korpus als Ganzes betreffen, eingeben (s. Kapitel *Korpusverwaltung*).
- **Daten**-Reiter wählt den zentrale Bereich des Programms an. Hier lassen sich die Metadaten zu Gesprächsereignissen, Konstellationen, Sprechern, Transkriptionen und Aufnahmen eingeben, filtern und selektieren (s. Kapitel *Verwaltung von Metadaten*).
- **Korpus-Korb**-Reiter bietet Zugriff auf einen „**Korb**“, in dem sich Transkriptionen zur späteren Untersuchung ablegen lassen. Das dahinter liegende Prinzip ist in der Einleitung kurz und im Kapitel *Korpus-Korb* näher beschrieben.

3.2 Menüs

In den Programmmenüs finden sich weitere Funktionen, die sich nicht direkt innerhalb des Programmfensters anwählen lassen:

3.2.1 Datei-Menü

Neues Dokument	Strg+N
Dokument öffnen...	Strg+O
Dokument speichern	Strg+S
Dokument speichern als...	
EXMARaLDA_DemoKorpus.coma	
^^ Liste löschen ^^	
Korb speichern als...	
Templates öffnen...	
Templates speichern	
Templates speichern als...	
Templates beim Start laden:	
Coma-Datei hinzufügen...	
Basistranskription importieren...	
Sprecher importieren...	
Korpus aus Transkriptionen erstellen	
Ausgabe...	
Programm beenden	Strg+Q

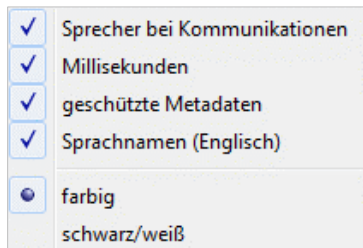
- **Neues Dokument:** Erstellt ein leeres Coma-Dokument und legt ein unbenanntes Korpus in ihm an. Nähere Informationen zu Korpora in Coma finden sie im Kapitel **Korpusverwaltung**.
- **Dokument öffnen...:** Zeigt ein Dateiauswahlfenster, über das sich eine Coma-Datei mit Metadaten öffnen lässt.
- **Dokument speichern:** Speichert die aktuell geöffnete Coma-Datei. Wenn die Datei noch nicht gespeichert wurde erscheint ein Dateiauswahlfenster, in dem sie den Dateinamen- und Ort angeben können. Beachten Sie dabei die Ausführungen unter **Korpusverwaltung** für einen sinnvollen Speicherort.
- **Dokument speichern als...:** Lässt Sie das Dokument an einem anderen Ort als dem aktuellen Speicherort speichern. Darauf folgt eine Liste der zuletzt geöffneten Dateien, die sich direkt anwählen lassen.
- **Liste löschen:** Löscht die Liste der zuletzt geöffneten Dateien.
- **Korb speichern als...:** Speichert den Korb (**Korpus-Korb**) als separates Korpus (s. auch **Korpus-Korb**).
- **Templates öffnen...:** Zeigt ein Dateiauswahlfenster, mit dem Sie eine Template-Datei öffnen können. Nähere Informationen zu Templates bekommen Sie unter **Templates**.
- **Templates speichern:** Speichert die aktuell geöffnete Templates-Datei mit allen angelegten Templates. Sollte keine Templates-Datei geöffnet sein, so erscheint ein Dateiauswahl-Dialog, in dem sich ein Speicherort- und Name für die Datei wählen lässt.
- **Templates speichern als...:** Speichert die angelegten Templates in eine Datei, die über einen Dateiauswahl-Dialog bestimmt wird.
- **Templates beim Start laden:** Bestimmt, ob die geladene Templates-Datei beim Start von Coma automatisch geöffnet wird.
- **Coma-Datei hinzufügen...:** Mit dieser Funktionen können andere existierende Coma-Dateien ausgewählt und in das geöffnete Dokument eingebunden („gemerged“) werden. Es werden sowohl die Kommunikationen als auch die Sprecher eingefügt.
- **Basistranskription importieren...:** Lässt sie Metadaten aus einer existierenden Basistranskription importieren (s. Basistranskriptionen Importieren).
- **Sprecher importieren:** Lässt Sie Sprecher-Metadaten aus einer anderen Coma-Datei importieren. Siehe (s. Importieren von Sprechern).
- **Korpus aus Transkriptionen erstellen:** Mit dieser Funktion lässt sich aus EXMARaLDA-Transkriptionen, die bereits mit Kopfdaten versehen sind, eine Coma-Datei erstellen. Diese Funktion ist unter **Korpus aus Transkriptionen erstellen** beschrieben.
- **Ausgabe...:** Erstellt eine .html-Ausgabedatei mit Metadaten, deren Pfad ausgewählt werden kann (s. **Ausgabe als html-Datei**).
- **Programm beenden:** Beendet das Programm.

3.2.2 Bearbeiten-Menü

Hier können die Einstellungen des Programms verändert werden. Siehe hierzu das Kapitel *Einstellungen*.

3.2.3 Ansicht-Menü

Die Optionen des **Ansicht**-Menüs wirken sich nur auf die Ansicht der Metadaten im Daten-Reiter (s. *Aufbau des Daten-Reiters*) aus.



- **Sprecher bei Kommunikationen**: Bestimmt, ob unter den Metadaten zu einer ausgewählten Kommunikation die Metadaten der zugeordneten Sprecher angezeigt werden sollen.
- **Millisekunden**: Bestimmt, ob bei Angaben zur „**Location**“ bei dem Schlüssel „**PeriodDuration**“ die Zeitspanne auch in Millisekunden angezeigt wird (sowohl bei Kommunikationen als auch bei Sprechern).
- **geschützte Metadaten**: Einige Metadatenfelder (wie die Anzahl der Segmente in einer segmentierten Transkription, s. Graphik unten) lassen sich in Coma nicht editieren. An dieser Stelle können Sie bestimmen, ob diese Felder angezeigt werden sollen oder nicht.

Segmented Transkription: Rudi_Voeller_Wutausbruch	
Description (Transcription)	
HIAT:ip	750
HIAT:non-pho	46
HIAT:u	68
HIAT:w	551
e	156
sc	36
segmented	true
File:	(...)er_Wutausbruch_s.exs

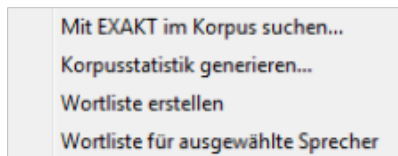
- **Sprachnamen (Englisch)**: Bestimmt, ob die Sprachkürzel in der Metadatenübersicht aufgelöst werden sollen. Angezeigt werden immer nur die Englischen Sprachnamen.
- **farbig, schwarz/weiß**: Bestimmt, ob die Metadatenansicht farbig oder monochrom erfolgt.

3.2.4 Werkzeuge-Menü

Hier kann das Werkzeug TreeTagger (siehe <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>) angewendet werden. Zur genaueren Erklärung der Funktionalität siehe *TreeTagger*.

3.2.5 Analyse-Menü

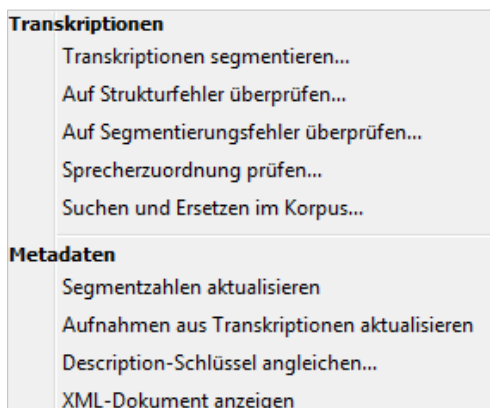
Das **Analyse**-Menü bietet Optionen, mit denen sich das Korpus als Ganzes analysieren lässt.



- **Mit EXAKT im Korpus suchen...**: Öffnet ein Suchfenster, mit dem sich in den Transkriptionen des geöffneten Korpus suchen lässt. Nähere Informationen entnehmen Sie bitte der **Dokumentation des Suchwerkzeuges EXAKT** (<http://exmaralda.org/> im Menü „Hilfe/Support“).
- **Korpusstatistik generieren...**: Ermöglicht es, eine Statistik über die im geöffneten Korpus vorhandenen Transkriptionen zu generieren (s. *Korpusstatistik erstellen*).
- **Wortliste erstellen**: Erstellt eine Wortliste der in der Transkription verwendeten Wörter (s. *Wortliste erstellen*). Diese Wortliste kann auch gefiltert oder als Datei gespeichert werden.
- **Wortliste für ausgewählte Sprecher**: Erstellt eine Wortliste der in der Transkription verwendeten Wörter für die Sprecher, die vorher markiert wurden (s. *Wortliste für ausgewählte Sprecher erstellen*).

3.2.6 Wartung-Menü

Im **Wartung**-Menü finden sich Optionen, mit denen sich Fehler in Metadaten sowie Transkriptionen des geöffneten Korpus bequem finden und beheben lassen. Das Menü ist in Optionen zu „**Transkriptionen**“ und zu „**Metadaten**“ unterteilt.



Transkriptionen:

- **Transkriptionen segmentieren**: Segmentiert die eingebundenen Basistranskriptionen nach einem gewählten Segmentieralgorithmus. Nähere Informationen zu verschiedenen Segmentierungsalgorithmen bekommen Sie in der Dokumentation zum EXMARaLDA-Partitur-Editor (s. *Transkriptionen segmentieren*).
- **Auf Strukturfehler überprüfen...**: Überprüft alle verknüpften Transkriptionen auf Strukturfehler (s. *Auf Strukturfehler überprüfen*).
- **Auf Segmentierungsfehler überprüfen...**: Überprüft alle verknüpften Transkriptionen auf Segmentierungsfehler (s. *Auf Segmentierungsfehler überprüfen*).

- **Sprecherzuordnung prüfen:** Überprüft, ob Sprecher im Korpus vorhanden sind, denen keine Transkriptionsspur zugeordnet ist (s. *Sprecherzuordnung prüfen*).
- **Suchen und Ersetzen im Korpus...:** Öffnet ein Menü, durch das die Basistranskriptionen des Korpus mit regulären Ausdrücken verändert werden können (s. *Suchen und Ersetzen im Korpus*).

Metadaten:

- **Segmentzahlen aktualisieren:** Extrahiert die Segmentzahlen aus allen verknüpften Transkriptionen und fügt diese in die Metadaten ein, so dass sie sich in Coma anzeigen und addieren lassen (s. *Segmentzahlen aktualisieren*).
- **Aufnahmen aus Transkriptionen aktualisieren:** Aktualisiert die Aufnahmen in Coma mit den Aufnahmen, die mit den Transkriptionen verknüpft sind. Bereits verknüpfte Aufnahmen, die nicht mehr existieren, werden dabei nicht gelöscht (s. *Aufnahmen aus Transkriptionen aktualisieren*).
- **Description-Schlüssel angleichen...:** Öffnet ein Fenster, in dem Sie Fehler, die Sie gegebenenfalls bei der Benennung von Metadaten-Schlüsseln gemacht haben, korrigieren können (s. *Description-Schlüssel angleichen*).
- **XML-Dokument anzeigen:** Öffnet ein Fenster mit der XML-Darstellung der Coma-Datei. Beachten Sie, dass es bei großen Coma-Dateien eine Weile dauern kann, bis das Fenster erscheint.


3.2.7 Hilfe-Menü


Im **Hilfe**-Menü finden sich Optionen zur Lösung von etwaigen Problemen mit Coma.




- **Hilfe (im Web):** Öffnet die Coma-Webseite mit Links zu diesem und weiteren Hilfe-Dokumenten.
- **Nach Updates suchen:** Stellt eine Verbindung mit dem EXMARaLDA-Webserver her und überprüft, ob eine neuere Version als die installierte von Coma erschienen ist. Falls dies der Fall ist, können Sie durch klicken auf die „**Download-Seite besuchen**“-Schaltfläche die EXMARaLDA-Download-Seite besuchen und die neueste Version herunterladen. In den *Einstellungen* können Sie bestimmen, ob die Update-Überprüfung in regelmäßigen Intervallen automatisch durchgeführt wird und ob *Preview-Versionen* in die Update-Prüfung mit einbezogen werden.
- **Debug-Information in die Zwischenablage kopieren:** Kopiert den Inhalt des Log-Files in die Zwischenablage. Wenn Sie einen Fehler im Programm entdecken, so können sie ein Email mit diesem Inhalt an die Programmierer senden.
- **Über Coma:** Zeigt ein Fenster mit Programminformationen und der Versionsnummer des Programms an.


4 Korpusverwaltung


Auf dem „**Korpus**“-Reiter werden alle Daten verwaltet, die sich mit dem Korpus als Ganzes befassen. Der Reiter: hier mit einem geladenen Korpus - sieht so aus:

 Dublin Core Metadata		Schlüssel	Wert
		DC:contributor	Hamburger Zentrum für Sprachkorpora, Max-Brauer-Allee 60 / ...
		DC:coverage	Covers mainly television and radio broadcasted speech records...
		DC:created	2007-11-08
		DC:creator	Hamburger Zentrum für Sprachkorpora, Max-Brauer-Allee 60 / ...
		DC:description	A linguistic corpus of spoken language (various languages, approx...
		DC:format	XML
		DC:identifier	http://www.exmaralda.org/en/korpora/demonstration/
		DC:language	DEU; ENG; FRA; SPA; TUR; POL; VIE; SWE; NOR; ITA; RUS; AF...
		DC:publisher	Hamburger Zentrum für Sprachkorpora, Max-Brauer-Allee 60 / ...
		DC:rights	Free to use for research and teaching purposes.
		DC:rightsHolder	Hamburger Zentrum für Sprachkorpora, Max-Brauer-Allee 60 / ...
		DC:subject	Linguistic corpus of spoken language
		DC:title	EXMARaLDA Demo corpus
		DC:type	Collection
		hzsk:corpusPrefix	demo
		hzsk:corpusVersion	1601
		hzsk:keywords	L1 data
		hzsk:shortDescription	A selection of short audio and video recordings in various langu...
		olac:compiler	Hamburger Zentrum für Sprachkorpora, Max-Brauer-Allee 60 / ...
		olac:data-inputter	Secil Yusun; Annette Schnieder; Andrea Rolle; Silke Merkel; Tho...
		olac:developer	Hamburger Zentrum für Sprachkorpora, Max-Brauer-Allee 60 / ...
		olac:researcher	Thomas Schmidt; Kai Wörner; Hanna Hedeland
		olac:sponsor	Deutsche Forschungsgemeinschaft (DFG)

Zu allen angelegten Korpora lassen sich in Coma beschreibende Metadaten eingeben: Vorgegeben sind dabei der „**Korpusname**“ und das Feld der „**Unique Speaker Distinction**“. Sie lassen sich im oberen Bereich des Programms editieren, durch einen Klick auf das folgende Symbol .

 Korpus
  Daten
  Korpus-Korb (0)

Korpusname
 **DEMO-KORPUS**

unique speaker distinction
 //speaker/abbreviation

Das Feld „**Unique Speaker Distinction**“ muss vom Anwender nur geändert werden, wenn die Sprecherabkürzungen für die erfassten Sprecher nicht eindeutig sind und die eindeutige Zuordnung über ein anderes Metadatenfeld erfolgt. In diesem Fall muss der XPath zu diesem Metadatenfeld an dieser Stelle eingegeben werden. Für Korpora, die aus bestehenden Transkriptionen erzeugt werden, wird dieser Wert automatisch generiert.

Zu jedem Korpus lässt sich in der großen Tabelle eine „**Description**“ (siehe *Die Coma-Datenstruktur*) anlegen, in der sich freie Attribut-Wert-Paare zu Eigenschaften des Korpus eingeben lassen. Die Schaltfläche „**Dublin Core Metadata**“ auf der linken Seite der Tabelle fügt die Schlüssel für eine Dublin-Core¹-konforme Ressourcenbeschreibung in die Tabelle ein - eine solche Beschreibung erleichtert die Auffindbarkeit eines Korpus z.B. im Internet. In der Abbildung sind für das ausgewählte Korpus Dublin-Core-Metadaten erfasst.

¹ See: <http://dublincore.org/>.

4.1 Korpus aus Transkriptionen erstellen

Die Funktion **Datei > Korpus aus Transkriptionen erstellen**, dient dazu, aus vorhandenen EXMARaLDA-Transkriptionen, bei denen bereits Metadaten als Kopfdaten eingegeben wurden, automatisch eine Coma-Datei zu erstellen.

Bevor man diese Funktion benutzt sollte man sich noch einmal klar machen, wie die Dateien eines EXMARaLDA-Korpus sinnvollerweise organisiert sein sollten:

- Mediendateien sollten im selben oder einem untergeordneten Verzeichnis liegen wie die Transkription, der sie zugeordnet sind;
- Transkriptionen, die thematisch oder organisatorisch zusammengehören sollten in Verzeichnissen zusammengefasst sein;
- Die Coma-Datei sollte oberhalb dieser Verzeichnisse gespeichert sein.

Wenn sie die „**Korpus aus Transkriptionen erstellen**“-Funktion verwenden, müssen Sie sich um den letzten Punkt nicht kümmern - die Datei wird von dem Assistenten an die richtige Stelle gespeichert.

4.1.1 Verwenden des Assistenten

Die Funktion zum Erstellen einer Coma-Datei aus Transkriptionen ist als Assistent aufgebaut, der sie durch die notwendigen Schritte führt.

Wenn Sie den Assistenten aufrufen sollten Sie folgenden Bildschirm sehen:

Korpus aus Transkriptionen erstellen

1. Coma-Datei
2. Transkriptionen
3. Segmentierung
4. Metadaten-Zuordnung
5. Personen
6. Zusammenfassung

Korpus-Name, Coma-Datei

Korpus-Name: NUAS

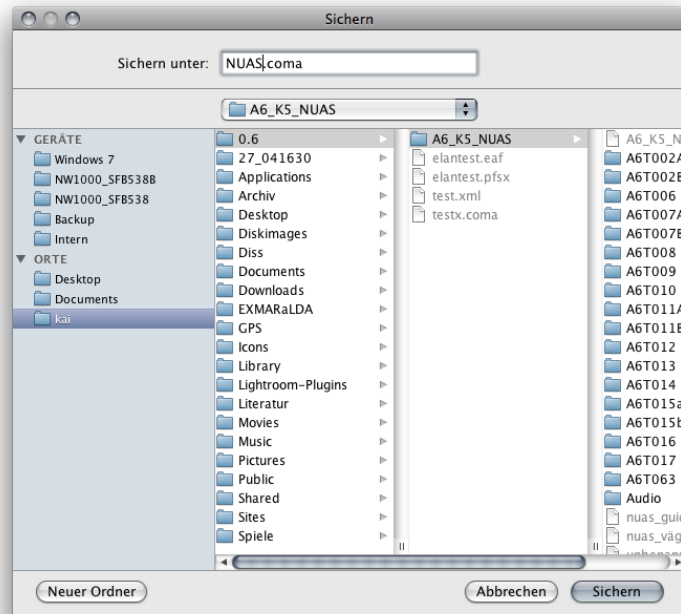
Coma-Datei: /Users/kai/0.6/A6_K5_NUAS/NUAS.coma

38 Transcriptions

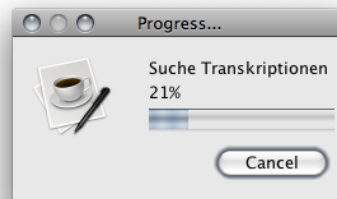
abbrechen < zurück weiter > abschliessen

Im Feld „**Korpus-Name**“ geben Sie den Namen ein, den Ihr Korpus später haben soll.

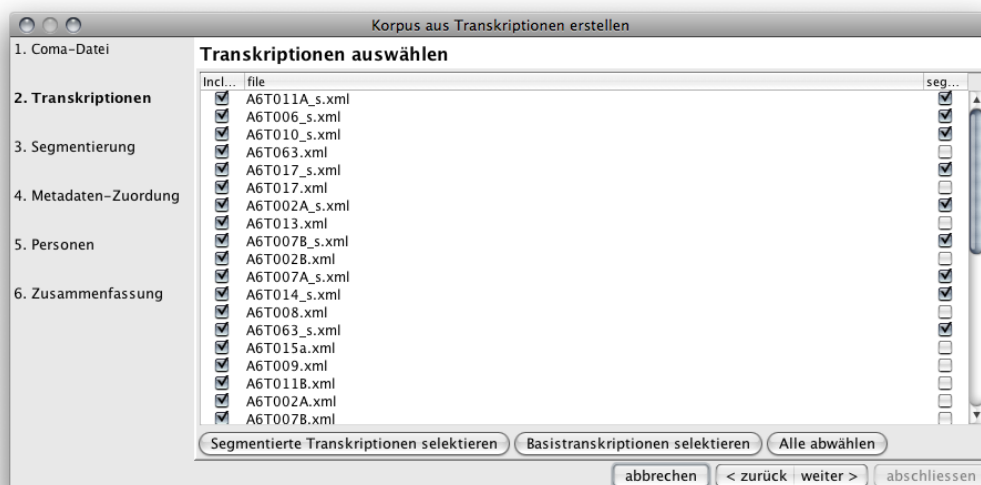
Klicken Sie dann auf die drei Punkte hinter dem Feld „**Coma-Datei**“ und wählen dort das Verzeichnis über dem Verzeichnis/den Verzeichnissen aus, in denen Ihre Transkriptionen liegen und wählen Sie ggf. einen anderen Dateinamen.



Daraufhin durchsucht der Assistent alle Verzeichnisse unterhalb des gewählten Verzeichnisses nach Transkriptionen und analysiert deren Kopf-/Metadaten.



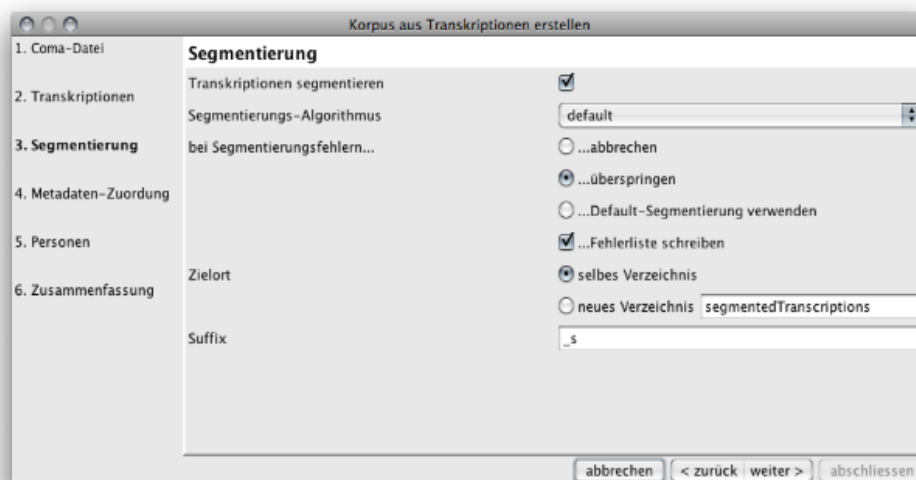
Klicken Sie daraufhin auf die **Weiter**-Schaltfläche, um zum Bildschirm mit der Auswahl der Transkriptionen zu kommen. Eine Zusammenfassung wird angezeigt.



In diesem Schritt können Sie festlegen, welche der gefundenen Transkriptionen in die Coma-Datei übernommen werden sollen. In der ersten Spalte der Tabelle können Sie Dateien an- oder abwählen, in der zweiten Spalte ist der Dateiname der Transkription angegeben und in der dritten Spalte wird angezeigt, ob es sich um eine segmentierte Transkription handelt oder nicht. Wollen Sie beispielsweise nur segmentierte Transkriptionen in ihre Coma-Datei übernehmen, so wählen sie zunächst mit der Schaltfläche **„alle abwählen“** am unteren Rand alle ab und fügen die segmentierten anschließend mit **„Segmentierte Transkriptionen wählen“** wieder hinzu.

Beachten Sie, dass die Auswahl der Transkriptionen natürlich einen Einfluss auf die Menge der Metadatenfelder haben kann, die Ihnen zur Verfügung stehen.

Wenn Sie Ihre Auswahl getroffen haben, klicken Sie auf **Weiter**.



Im nächsten Bildschirm können Sie wählen, ob Basistranskriptionen automatisch segmentiert werden sollen. Nähere Informationen zu verschiedenen Segmentierungsalgorithmen bekommen Sie in der **Dokumentation zum EXMARaLDA-Partitur-Editor** (www.exmaralda.org, Sektion „Hilfe/Support“)

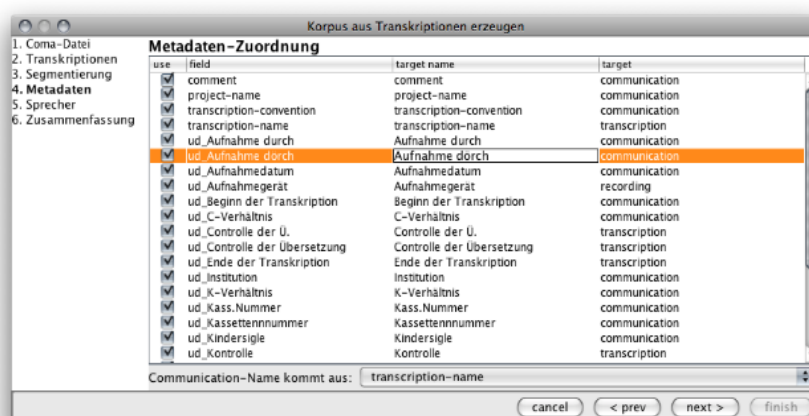
- Das Kontrollfeld **„Transkriptionen segmentieren“** bestimmt, ob überhaupt Basistranskriptionen segmentiert werden sollen. Wenn dieses Kontrollfeld abgewählt ist, stehen auch die anderen Optionen nicht zur Verfügung.
- Beim Auswahlmenü **„Segmentierungs-Algorithmus“** können Sie zwischen den verfügbaren Algorithmen auswählen.

Die nächsten Kontrollfelder bestimmen, was die Segmentierungsroutine machen soll, wenn sich eine Transkription aufgrund eines Fehlers (meist bedingt durch nicht eingehaltene Transkriptionskonventionen) nicht segmentieren lässt:

- Bei gewählter Option **„abbrechen“** wird der Segmentierungsprozess abgebrochen, sobald sich eine Transkription nicht segmentieren lässt. Auch die vorher bereits segmentierten Transkriptionen werden wieder gelöscht.
- Wenn die **„überspringen“**-Option angewählt ist, werden nur die fehlerhaften Transkriptionen nicht segmentiert, die Transkriptionen, bei denen keine Fehler auftreten, werden dagegen geschrieben.

- Mit der Option „**Default-Segmentierung verwenden**“ werden Transkriptionen, bei denen ein Fehler auftritt, mithilfe des Default-Algorithmus behandelt, der zwar keine eigentliche Segmentierung durchführt, aber trotzdem eine Transkription vom Typ „segmentierte Transkription“ erstellt.
- Mit der Checkbox „**Fehlerliste schreiben**“ können Sie bestimmen, ob eine Liste mit den Fehlern, die bei der Segmentierung aufgetreten sind, geschrieben wird. Die Datei wird an dieselbe Stelle geschrieben wie die Coma-Datei (die Stelle, die am Anfang ausgewählt wurde) und heißt „segmentation-errors.xml“. Diese Datei können Sie benutzen, um die Segmentierungsfehler direkt im Partitureditor zu korrigieren (*Fehlerliste*).
- Unter „**Zielort**“ können Sie bestimmen, an welche Stelle die segmentierten Transkriptionen geschrieben werden sollen – entweder in dasselbe Verzeichnis wie die Basistranskriptionen („**selbes Verzeichnis**“) oder in ein „**neues Verzeichnis**“ im Stammordner des Korpus. Den Namen des neuen Verzeichnisses müssen Sie in dem Eingabefeld bestimmen.
- Mit „**Suffix**“ bestimmen Sie, was an den Dateinamen der Transkriptionen angehängt wird, um zu kennzeichnen, dass es sich um segmentierte Transkriptionen handelt.

Wenn Sie die Einstellungen vorgenommen haben, klicken Sie auf **weiter**.



Auf dem Bildschirm „**Metadaten-Zuordnung**“ bestimmen Sie, welche der gefundenen Metadaten mit welchem Namen und an welche Stelle in der Coma-Datei übernommen werden.

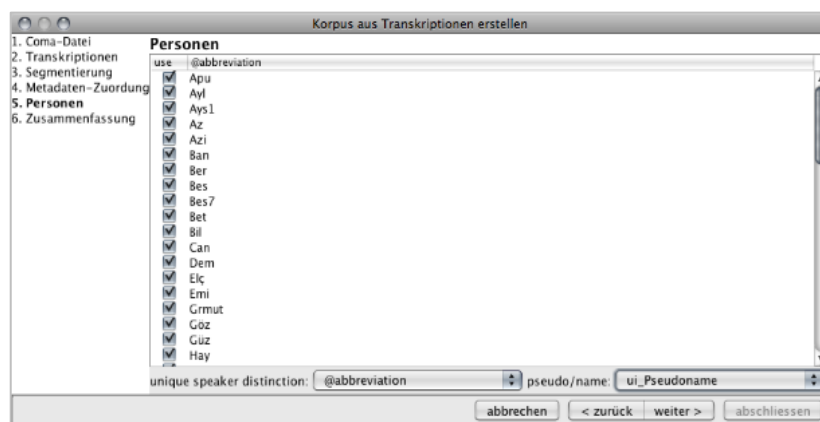
- In der ersten Spalte können Sie bestimmen, ob der jeweilige Schlüssel überhaupt in die Coma-Datei übernommen werden soll oder nicht.
- In der zweiten Spalte sehen Sie alle Transkriptions-Kopfdatenschlüssel (außer denen der Sprechertabellen), die in allen Transkriptionen des Korpus gefunden wurden. Felder, die als „frei bestimmbare Attribute“ eingegeben wurden haben in der Liste ein vorangestelltes „**ud_**“ (für „user defined“).
- In der folgenden dritten Spalte („**target name**“) können Sie den Namen bestimmen, die die entsprechenden Schlüssel in Coma bekommen sollen; in der Regel werden das die vorgegebenen sein, die bereits im Partitur-Editor vergeben wurden. An dieser Stelle lassen sich allerdings auch Daten vereinheitlichen: Im obigen Beispiel ist der Schlüssel „**ud_Aufnahme_dörch**“ markiert, der direkt unterhalb des Schlüssels

„**ud_Aufnahme_durch**“ steht - offensichtlich ein Eingabefehler. In der dritten Spalte liesse sich an dieser Stelle „**Aufnahme durch**“ eintragen, um die Felder in Coma zu vereinheitlichen.

- In der vierten Spalte „**target**“ („Ziel“) wird bestimmt, an welcher Stelle in der Coma-Datei die Metadaten schließlich landen: Metadaten, die in einem Transkriptionskopf standen beziehen sich ja nicht notwendigerweise auf eine Coma-Communication. Zur Auswahl stehen „**communication**“ für Metadaten zur Gesprächskonstellation, „**recording**“ für Metadaten zur Aufnahme und „**transcription**“ für Metadaten zur Transkription.
- Häufig existieren zu einer Gesprächssituation („communication“) mehrere Transkriptionen. Im Drop-Down-Menü am unteren Bildschirmrand bestimmen Sie, nach welchen Metadaten-schlüssel Transkriptionen zu einer Kommunikation zusammengefasst werden - wenn sie also mehrere Transkriptionen in Kommunikationen zusammenfassen wollen, sollten Sie in den Transkriptionen einen Metadaten-schlüssel haben, der die entsprechende Kommunikation eindeutig definiert. Wenn jede Transkription genau einer Kommunikation zugeordnet ist, dann wählen Sie „**one file -> one communication**“ aus dem Menü.

Sollten Sie nach dem Schreiben der Coma-Datei feststellen, dass sie an dieser Stelle irgendetwas falsch zugeordnet haben, so können Sie einfach mit dem **zurück**-Knopf wieder an diese Stelle springen und die Zuordnungen ändern - der eigentliche Vorgang des Erstellens der Coma-Datei geht nämlich relativ schnell vonstatten.

Wenn Sie die Zuordnungen vorgenommen haben, klicken Sie auf **weiter**.

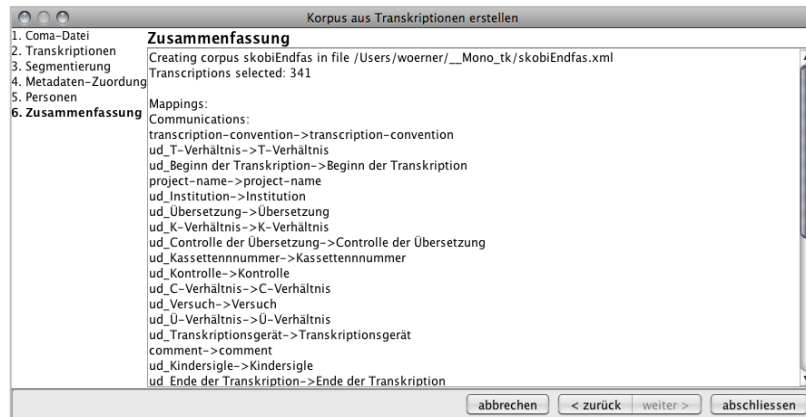


Im nächsten Bildschirm können Sie bestimmen, welche der gefundenen Sprecher („**Personen**“) in die Coma-Datei übernommen werden soll, indem sie in der ersten Spalte der Tabelle die Sprecher an- oder abwählen.

- Mit dem linken Drop-Down-Menü bestimmen Sie, anhand welches Metadatenfeldes in den Partitur-Sprechertabellen die jeweiligen Sprecher im Korpus auseinandergehalten werden. In den meisten Fällen wird das Namenskürzel aus der im Folgenden abgebildeten Sprechertabelle im Partitur-Editor für die jeweiligen Sprecher eindeutig sein, in den wenigsten Fällen die durchnummerierten Sprecherkürzel („SPK0, SPK1...“), die im Assistenten als „Id“ erscheinen. Wenn Sie die Auswahl im linken Drop-Down-Menü ändern, so wird die Tabelle mit den zu importierenden Sprechern sofort aktualisiert - so können Sie gleich erkennen, ob Sie das richtige Feld für eine eindeutige Zuordnung ausgewählt haben.

- In Coma haben alle Sprecher ein Kürzel und einen (Pseudo-)Namen. Aus welchem Metadatenfeld dieser Name bestimmt werden soll, bestimmen Sie mit dem rechten Drop-Down-Menü.

Wenn Sie die entsprechenden Auswahlen vorgenommen haben, klicken Sie auf **weiter**.



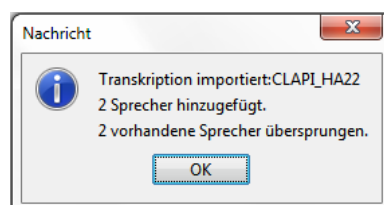
Im nächsten Bildschirm sehen Sie noch einmal eine Zusammenfassung der vorgenommenen Einstellungen. Wenn Sie noch etwas ändern wollen, klicken Sie auf die Schaltfläche **zurück** und nehmen Sie Änderungen vor. Ansonsten klicken Sie auf **abschliessen**, um die Coma-Datei zu erzeugen. Wenn die Datei fehlerlos geschrieben wurde wird das Fenster geschlossen und die Coma-Datei in Coma geöffnet.

4.2 Importieren und Exportieren von Metadaten

4.2.1 Basistranskriptionen Importieren

Mit dieser Option lassen sich Metadaten aus einer EXMARaLDA-Basistranskription importieren. Wenn Sie den Menüpunkt anwählen, erscheint ein Dateiauswahlfenster, in dem Sie die Basistranskription auswählen, die Sie in Coma integrieren wollen. Beachten Sie auch hier wieder, dass die Datei in einem Verzeichnis unterhalb der Coma-Datei (oder im selben Verzeichnis) liegen muss.

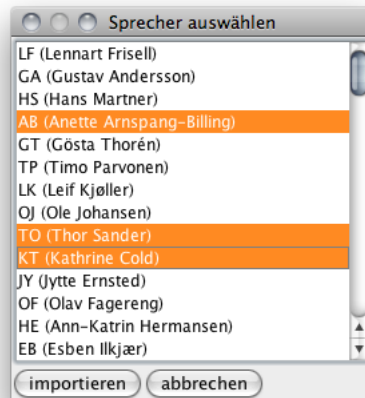
Weitere Eingaben sind nicht erforderlich; sofern die Transkription noch nicht mit der Coma-Datei verknüpft ist, wird eine neue Kommunikation und neue Sprecher (sofern noch nicht vorhanden) mit den Metadaten der Transkription angelegt und diese gleichzeitig verknüpft.



4.2.2 Sprecher importieren

Um Sprecher, die in einer anderen Coma-Datei bereits angelegt sind, in die geöffnete Coma-Datei zu importieren, wählen Sie **Datei > Sprecher importieren**. Daraufhin öffnet sich ein Dateiauswahl-Fenster, in dem Sie die Coma-Datei auswählen, in der sich die Metadaten der

Sprecher befinden, die sie importieren wollen. Wenn Sie die Datei ausgewählt haben, öffnet sich ein Dialog mit einer Liste aller Sprecher, die in der ausgewählten Datei angelegt sind:



Wählen Sie aus der Liste die Sprecher aus, die Sie importieren wollen, und bestätigen Sie die Auswahl mit **importieren**. Die Metadaten der entsprechenden Sprecher werden daraufhin importiert.

5 Verwaltung von Metadaten

Alle Metadaten zu Gesprächskonstellationen, Sprechern, Aufnahmen und Transkriptionen werden auf dem „**Daten**“-Reiter vorgenommen.

5.1 Aufbau des Daten-Reiters

Der Bildschirm gliedert sich in drei Bereiche:

- Auf der linken Seite befindet sich eine Tabelle mit den Kommunikationen des gewählten Korpus. Über dieser Tabelle liegt der Bereich, in dem sich Filter für diese Tabelle modifizieren und anlegen lassen (**Filter**). Unter der Tabelle finden sich Schaltflächen zum Hinzufügen, Entfernen und Duplizieren von Kommunikationen sowie für die Bedienung des **Korpus-Korbes** und zur **Segment-Zählung**:




- In der mittleren Spalte werden die Metadaten der ausgewählten Kommunikationen oder Sprecher angezeigt und bearbeitet. Über dieser Anzeige findet sich ein Suchfeld, mit dem sich im Volltext der Metadaten **Suchen** lässt:

- Unterhalb der Anzeige findet sich eine Schaltfläche zum Drucken der **Metadatenansicht**, sowie zwei Schaltflächen zum **Verknüpfen von Kommunikationen und Sprechern**.

- In der rechten Spalte werden – analog zur Kommunikationen-Spalte – die Sprecher verwaltet.
- Die Trenner zwischen den drei Spalten lassen sich verschieben, um mehr Platz für einzelne Spalten auf dem Bildschirm zu bekommen. Am oberen Rand dieser Trenner befinden sich kleine Dreiecke - mit einem Klick auf diese Pfeile lassen sich Spalten ganz ein- oder ausblenden, so dass sich beispielsweise nur Kommunikationen und deren Metadaten anzeigen lassen.

5.2 Verwalten von Kommunikationen

Mit den Kommunikationen werden alle Konstellationsdaten zu dem Gesprächsereignis gespeichert, also Dinge wie Ort, Zeit, besondere Umstände, beteiligte Sprachen etc.

Kommunikation		
S	Name	Kommunikationsname
	A6D126	Abschluss Tusby
	A6D057	Pause zwischen Vorträgen
	A6D120	Diskussion nach Mertanos Vortrag
	A6D084	Diskussionsbeiträge/Kommentar...
	A6D105	Podiumsdiskussion Århus 2
	A6D114	Podiumsdiskussion: "Elektroniske..."

Die „**Kommunikationen**“ werden in einer dreispaltigen Tabelle auf der linken Seite des Reiters angezeigt. In der zweiten Spalte steht der „**Kommunikationsname**“, ein in Coma vorgegebenes Feld, in der dritten Spalte erscheint ein wählbares Element aus den Metadaten zu der jeweiligen Kommunikation – im Falle des obigen Screenshots ein ausführlicherer Name für die Kommunikation. Das Element, das hier angezeigt wird, wird durch einen Klick auf den entsprechenden Schlüssel der Kommunikationen oder Sprecher gewählt.





In der ersten Spalte erscheint ein Büroklammer-Symbol, falls diese Kommunikation mit einem in der Sprecher-Spalte ausgewählten Sprecher verknüpft ist. Wenn auf der rechten Seite ein oder mehrere Sprecher angewählt werden kann man so erkennen, ob diese Sprecher der jeweiligen Kommunikation zugeordnet sind, d.h. an ihr teilgenommen haben.

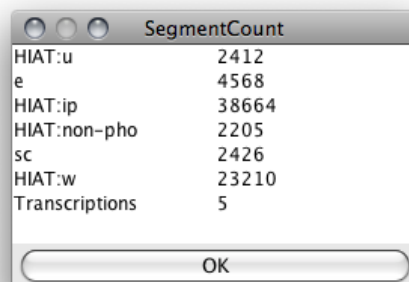
Mit einem Klick auf die Spaltenüberschriften lässt sich die Tabelle nach dieser Spalte sortieren - dies ist insbesondere für die Warenkorb-Funktionalität hilfreich, aber auch, um sich beispielsweise schnell einen Überblick zu verschaffen, an welchen Kommunikationen ein bestimmter Sprecher teilgenommen hat: Sortieren Sie dazu die Tabelle anhand der ersten Spalte und wählen Sie auf der rechten Seite einen Sprecher aus - die relevanten Kommunikationen stehen jetzt untereinander am Anfang (oder am Ende) der Tabelle.

Mit den Schaltflächen unterhalb der Tabelle lassen sich u.a. Kommunikationen anlegen und löschen.



- Ein Klick auf diese Schaltfläche legt eine neue Kommunikation an. Sie erscheint üblicherweise am Ende der Tabelle und hat den Namen „**unnamed**“. Eine eindeutige ID wird automatisch vergeben und kann in Coma nicht geändert werden.

-  Diese Schaltfläche löscht die ausgewählte(n) Kommunikationen. Achtung: Es erfolgt keine Sicherheitsabfrage und dieser Schritt lässt sich nicht rückgängig machen.
-  Diese Schaltfläche erstellt eine Kopie der ausgewählten Transkription. Die Kopie erhält eine neue Id und den Namen der Quell-Kommunikation mit einem angehängten „_cloned“, die Metadaten sind ansonsten identisch. Sind beim Betätigen der Schaltfläche mehrere Kommunikationen ausgewählt, so wird nur von der obersten eine Kopie erstellt.
-  Mit dieser Schaltfläche werden alle Transkriptionen, die den augenblicklich ausgewählten Kommunikationen zugeordnet sind, in den „**Korpus-Korb**“ gelegt. Nähere Informationen finden sich im Kapitel zum **Korpus-Korb**.
-  Die Taschenrechner-Schaltfläche führt eine Segment-Zählung durch. Dazu addiert sie die Segmentzahlen aller segmentierten Transkriptionen, die den ausgewählten Kommunikationen zugeordnet sind, und gibt sie in einem kleinen Fenster aus. In diesem Beispiel:



HIAT:u	2412
e	4568
HIAT:ip	38664
HIAT:non-pho	2205
sc	2426
HIAT:w	23210
Transcriptions	5

OK

haben die (nach HIAT) segmentierten Transkriptionen zu den ausgewählten Kommunikationen beispielsweise 2412 *utterances* und 23210 *words*.

5.3 Metadaten-Anzeige

Wenn eine Kommunikation ausgewählt ist, werden in der mittleren Spalte die erfassten Metadaten angezeigt.

Im oberen Balken steht der Name der Kommunikation, der auch immer in der zweiten Spalte der Kommunikations-Tabelle erscheint. Es folgen die „**Descriptions**“ zur Kommunikation und weitere Metadatenfelder, die ebenfalls über eine Description verfügen können.

In der Metadaten-Ansicht lassen sich einige Elemente anklicken, um mit dem Programm zu interagieren:

- Mit einem Klick auf den Metadaten Schlüssel werden die Werte des Metadaten Schlüssels, der sich in dieser Zeile befindet, als zweite Spalte in der Tabelle mit den Kommunikationen angezeigt. Möchte man sich also z.B. einen schnellen Überblick darüber verschaffen, zu welchem Gesprächstyp die jeweiligen Kommunikationen im obigen Beispiel gehören, so würde man auf den Schlüsselnamen „Gesprächstyp“ in der Metadatenansicht klicken.
- Mit einem Klick auf einen Metadatenwert wird ein Filter für die Kommunikationstabelle erstellt. Klickt man im obigen Beispiel auf den Wert „Fernsehdebatte“ hinter dem Metadaten Schlüssel „Gesprächstyp“, so werden in der Kommunikationstabelle nur noch Kommunikationen angezeigt, bei denen dem Attribut „Gesprächstyp“ der Wert „Fernsehdebatte“ zugewiesen ist. Nähere Informationen zu der Verwendung von Filtern finden sie unter **Filter**.

Metadaten

Suchen [Suchen] [reset]

Communication Nguyen Ngoc Ngan

Description (Communication)

Background information	Interview mit vietnamesisch-kanadischem Schriftsteller Nguyen Ngoc Ngan
Communication type	television interview
Project name	EXMARaLDA DemoKorpus

Communication (Location)

Country	Vietnam
PeriodStart	19.01.2009 00:00:00

Description (Location)

Languages

Communication (Language)

LanguageCode	vie
--------------	-----

Description (Language)

Setting

Description (Setting)

1 Recording

Recording: NguyenNgocNgan (00:05:48.507; 348507ms)

Description (Recording)

File	NguyenNgocNgan.mp3
File	NguyenNgocNgan.wav
File	NguyenNgocNgan.ogg

2 Transcriptions

Basic Transkription: NguyenNgocNgan




Description (Transcription)

segmented	false
File	(...)/n/NguyenNgocNgan.exb

Segmented Transkription: NguyenNgocNgan

Description (Transcription)

Alignment status	fully aligned
Annotation type: en	English translation
Segmentation algorithm	HIAT

-  Mit dieser Schaltfläche lassen sich die Metadaten editieren. Ein Klick auf das Symbol öffnet je nach Metadatenfeld unterschiedliche Dialoge, in denen sich die entsprechenden Änderungen vornehmen lassen. Die einzelnen **Eingabedialoge** werden ein wenig später erklärt.
-  Analog zu den anderen Schaltflächen in Coma lassen sich mit den Plus- und Minus-Schaltflächen Metadatenfelder hinzufügen oder entfernen. Im obigen Beispiel ist ein Plus-Symbol neben der Überschrift „**Language(s)**“; mit einem Klick auf das Symbol ließe sich der Kommunikation eine Sprache zuordnen.
-  Mit einem Klick auf das Minus-Symbol neben der „**Recording**“ ließe sich diese bereits zugeordnete Aufnahme nach einer Sicherheitsabfrage entfernen.

Es werden keine Metadaten angezeigt, wenn mehrere Kommunikationen ausgewählt sind. In diesem Fall wird nur die Anzahl ausgewählter Kommunikationen und zugeordneter Sprecher angezeigt:

**13 Kommunikationen ausgewählt,
140 zugeordnete Sprecher**

5.4 Eingabedialoge



Wenn in der Metadatenanzeige auf diese Schaltfläche gedrückt wird, erscheint je nach ausgewähltem Metadatenfeld ein Dialog, mit dem dieses Metadatenfeld bearbeitet werden kann. Diese Dialogfelder werden im Folgenden kurz erläutert.

5.4.1 Location

Das Location-Dialogfeld dient zur Bearbeitung des Datentyps „**Location**“ – es handelt sich dabei um einen speziellen Datentyp, der bei **Locations** besprochen ist.

„**Locations**“ beschreiben – wie dort erwähnt – eine Kombination von Zeit(-abschnitt) und Ort.

Das Feld „**Type**“ kann eine nähere Beschreibung der Location enthalten. Wird hier ein Wert eingegeben, so erscheint in der Metadatendarstellung später statt der Bezeichnung „**Location**“ der entsprechende Wert.

Die Felder „**Street**“, „**City**“, „**PostalCode**“ und „**Country**“ erklären sich weitgehend von selber. Die (optionale) Startzeit lässt sich bequem durch Anklicken des Kalendersymbols an der rechten Seite einstellen: Es erscheint ein Kalender-Auswahlfenster, in dem das Datum nur angeklickt werden muss.

Die Dauer wird in dem Feld eingegeben, das aus mehreren kleinen Eingabefeldern besteht. Diese stehen (von rechts nach links) für: Jahre, Monate, Tage, Stunden, Minuten, Millisekunden. Will man also die Dauer von 2 Jahren und 3 Monaten eintragen, so genügt es, die ersten beiden Felder mit 2 und 3 auszufüllen und die restlichen Felder leer zu lassen.

Unterhalb findet sich das Feld zur Eingabe der „**Description**“, die die Location näher beschreibt.

Am oberen Rand des Eingabefensters findet sich ein Panel, mit dem sich Templates anlegen und abrufen lassen. Dies ist unter *Templates* näher besprochen.

5.4.2 Language

Der Language-Eingabedialog dient der Erfassung von Sprachen bei Sprechern und Kommunikationen. In Coma werden Sprachen in Form von 3-stelligen Sprachkürzeln erfasst, die dem „**ISO-Standard 639-3**“ entsprechen².

In das Feld „**ISO 639-3 Code**“ wird der Code für die gewünschte Sprache direkt eingetragen. Mit einem Druck auf die **Enter**-Taste wird der eingegeben Code in der Tabelle verfügbarer

² Siehe <http://www.ethnologue.com/codes/default.asp>

Codes nachgeschlagen und ein (englischer) Name für die Sprache mit diesem Kürzel erscheint in dem Feld „**Name**“. Erscheint dort kein Name, so existiert dieser Code vermutlich nicht.

Durch die Eingabe eines (englischen) Sprachnamens in das Feld „**Name**“ können Sie das gesuchte Kürzel ermitteln. Geben Sie dazu den Namen (oder einen Teil des Namens) der gesuchten Sprache ein und wählen Sie die Sprache aus den angezeigten Vorschlägen. Das entsprechende Sprachenkürzel wird dann in das ISO-Feld eingetragen. Der Inhalt des Feldes „**Name**“ selber dient nur der Orientierung und zur Unterstützung bei der Eingabe der Kürzel – der dort hinterlegte Wert wird in der Coma-Datei nicht gespeichert.

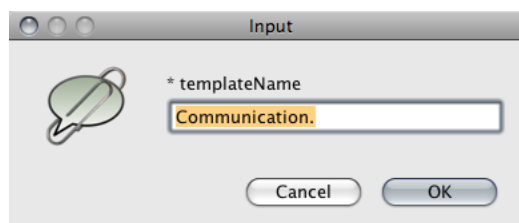
Unter „**Type**“ lässt sich die erfasste Sprache für den aktuellen Sprecher oder die aktuelle Kommunikation (analog zum „**Type**“ bei Locations) näher Kategorisieren. Für Sprecher bietet es sich an, hier den Status (L1, L2, ...) der Sprache für diesen Sprecher anzugeben – dieser wird dann in der Metadatenübersicht direkt angezeigt.

5.5 Templates

Für häufig benötigte Datensätze lassen sich in Coma „**Templates**“ anlegen. Die Datentypen, für die Templates zur Verfügung stehen, erkennt man daran, dass sich ein standardisierter Template-Balken am oberen Rand des Eingabefensters findet.



Mit einem Klick auf den Plus-Button werden die Daten, die im Eingabefenster eingegeben wurden, in ein neues Template übernommen. Es erscheint ein Eingabefenster, in dem sich ein Name für das Template eingeben lässt. Danach lassen sich Eingabefenster desselben Datentyps (also z.B. für „**Communications**“ oder „**Descriptions**“) automatisch mit diesen Werten ausfüllen, indem aus dem Dropdown-Menü das Template mit dem vergebenen Namen ausgewählt wird.



Ein Klick auf die Minus-Schaltfläche löscht das im Dropdown-Menü ausgewählte Template.

5.5.1 Templates speichern

Die angelegten Templates lassen sich über das Datei-Menü abspeichern und wieder öffnen. Dort (und in den *Einstellungen*) lässt sich auch festlegen, dass eine bestimmte Template-Datei beim Start des Corpus-Managers automatisch geöffnet wird (s. Graphik unten).

Neues Dokument	⌘N
Dokument öffnen...	⌘O
Dokument speichern	⌘S
Dokument speichern als...	
EXMARaLDA_DemoKorpus.xml	
test.xml	
Templates öffnen...	
Templates speichern	
Templates speichern als...	
✓ Templates beim Start laden: comatemplates.ctf	
Korpus aus Transkriptionen erstellen	
Programm beenden	

5.6 Verwalten von Personen und Sprechern

Die Verwaltung von Personen und Sprechern erfolgt grundsätzlich analog zu der Verwaltung von Kommunikationen.

Speaker: SR (SR, Sex: female)	
Description (Speaker)	
Beruf	Politikerin
Name	Royal
Vorname	Ségolène
Location	
Geburt (Location)	
City	Dakar
Country	Senegal
PeriodStart	22.09.1953 00:00:00
Description (Location)	
Language(s)	
L1	
LanguageCode	fra
Description (Language)	

Zu beachten ist, dass für Sprecher – im Gegensatz zu Kommunikationen – mehrere Locations angelegt werden können, um Stationen der Biographie von Sprechern abbilden zu können.

5.6.1 Importieren von Sprechern

Wenn Metadaten zu Sprechern bereits in einer anderen Coma-Datei vorliegen, so lassen sich diese in die geöffnete Coma-Datei importieren (*Importieren von Sprechern*).

5.7 Verknüpfen von Kommunikationen und Sprechern



Um „**Verknüpfungen**“ zwischen Kommunikationen und den beteiligten Sprechern herzustellen, markieren sie in der Kommunikationen-Tabelle die gewünschten Kommunikationen (auch mehrere - verwenden sie dazu die betriebssystemspezifischen Tasten zur Auswahl mehrerer Objekte) und in der Sprechertabelle die beteiligten Sprecher und klicken Sie auf die „**Verknüpfen**“-Schaltfläche (Kettensymbol).

Beachten Sie: Es werden alle selektierten Sprecher zu allen selektierten Kommunikationen hinzugefügt, unabhängig davon, ob sie bereits verknüpft waren oder nicht.



Um Verknüpfungen zwischen Sprechern und Kommunikationen aufzuheben, verfahren Sie analog: Markieren Sie alle Kommunikationen und Sprecher, zwischen denen Sie die Verknüpfung aufheben wollen, und klicken Sie auf die „**Verknüpfung aufheben**“-Schaltfläche.

Alle Verknüpfungen zwischen den gewählten Kommunikationen werden aufgehoben, ob sie vorher bestanden oder nicht.

ACHTUNG: Beide Aktionen können nicht rückgängig gemacht werden!

5.8 Verwalten von Transkriptionen

Alle Transkriptionen sind in Coma Kommunikationen zugeordnet und werden entsprechend mit ihnen verwaltet.

2 Transcriptions +	
Segmented Transkription: Rudi_Voeller_Wutausbruch	
Description (Transcription)	
HIAT:ip	750
HIAT:non-pho	46
HIAT:u	68
HIAT:w	551
e	156
sc	36
segmented	true
File:	(...)er_Wutausbruch_s.exs
Basic Transkription: Rudi_Voeller_Wutausbruch	
Description (Transcription)	
segmented	false
File:	(...)ller_Wutausbruch.exb

Wenn Sie sich die Metadaten zu einer Kommunikation anzeigen lassen, so sehen Sie immer einen Abschnitt für die Verwaltung von Transkriptionen.



Analog zu den anderen Metadatenobjekten können Sie mit einem Klick auf diese Schaltfläche eine Transkription hinzufügen. Es öffnet sich daraufhin ein Dateiauswahlfenster, mit dem Sie die Transkription, die zur ausgewählten Kommunikation gehört, auswählen.

Beachten Sie, dass die Transkriptionen im selben Verzeichnis wie die Coma-Datei oder in einem Verzeichnis unter dem Verzeichnis der Coma-Datei liegen müssen.

Wenn Sie die Datei(en) ausgewählt haben, wird die Transkription der ausgewählten Kommunikation zugeordnet. Die Transkription bekommt in Coma einen Namen – dieser entspricht entweder dem Dateinamen oder dem Namen der Kommunikation, der sie zugeordnet ist – welche Option gewählt wird lässt sich in den *Einstellungen* bestimmen. Bei mehreren Transkriptionen werden an die Namen fortlaufende Nummern angehängt.

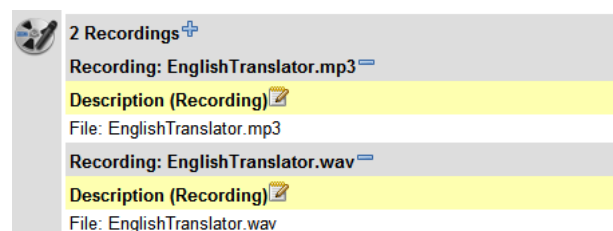
Bei der Zuordnung wird in die Transkriptionsdatei hineingeschaut. Dabei wird überprüft, ob es sich überhaupt um eine EXMARaLDA-Transkription handelt, ob es eine Segmentierte- oder eine Basistranskription ist und es werden die Metadaten zu Segmentzahlen ausgelesen. Es findet allerdings keine weitergehende Prüfung auf Korrektheit der Transkription statt. Die ermittelten Segmentzahlen werden in die Description der neu angelegten Transkription in Coma eingefügt, lassen sich aber nicht mit der Description editieren – es sind sogenannte „geschützte Metadaten“, deren Schlüsselname mit einem „#“-Zeichen beginnt. Im **Ansicht**-Menü lässt sich die Anzeige dieser Metadaten ein- oder ausschalten.

Die Transkriptionen bekommen auch eine eindeutige ID in Coma, die einer eindeutigen ID in der Transkriptionsdatei entsprechen sollte. Standardmäßig schreibt Coma diese ID bei der Zuordnung in den Kopf der Transkriptionsdatei hinein. Sollten Sie nicht wünschen, dass Coma in die Transkriptionsdateien schreibt, so können Sie das ebenfalls in den *Einstellungen* abschalten (d.h. „**Coma-ID beim Zuordnen in Transkriptionen schreiben (sofern nicht vorhanden)**“).

— Mit dieser Schaltfläche können Sie die Zuordnung der Transkription löschen. Beachten Sie, dass die eigentliche Transkription nicht gelöscht wird – dies müssen Sie von Hand erledigen.

5.9 Verwalten von Aufnahmen

Die Verwaltung von Aufnahmen erfolgt analog zu der von Transkriptionen direkt in der Metadatenansicht von Kommunikationen:



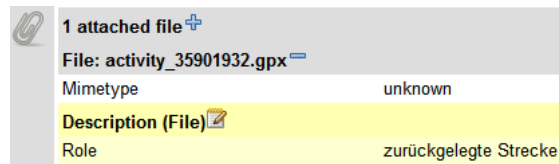
+ Mit dieser Schaltfläche ordnen Sie eine Aufnahme einer Kommunikation zu. Wenn Sie die Schaltfläche betätigen, erscheint ein Dateiauswahlfenster, in dem Sie die entsprechende(n) Mediendatei(en) auswählen können. Wenn sie die Auswahl bestätigt haben so wird eine Verknüpfung zu dieser Datei innerhalb der Kommunikation angelegt. Die Coma-Datei muss vorher gespeichert sein, da der Link zu der Mediendatei relativ zur gespeicherten Coma-Datei angelegt wird.

— Mit dieser Schaltfläche entfernen Sie eine Verknüpfung mit einer Aufnahme.

Sofern Java für den Dateityp einer Aufnahme einen entsprechenden Player vorgesehen hat, so können Sie mit einem Klick auf den Dateinamen, der hinter „**File:**“ steht, die Mediendatei direkt aus Coma abspielen.

5.10 Verwaltung von zusätzlichen Dateien

Zusätzlich zu Transkriptionen und Aufnahmen lassen sich an Kommunikationen auch Dateien anhängen, die nicht mit EXMARaLDA-Werkzeugen verarbeitet werden können. Diese können beispielsweise für die Situationsbeschreibung interessant sein.



Die Verwaltung dieser Dateien erfolgt analog zu der von Aufnahmen. Beachten Sie auch hier, dass die Dateien in einem Verzeichnis unterhalb der Coma-Datei liegen müssen, um die Auflösung der relativen Pfade gewährleisten zu können.

6 Analysieren von Daten

6.1 Filter

Mit Filtern lässt sich bestimmen, welche Kommunikationen in der linken und welche Sprecher in der rechten Spalte angezeigt werden.

6.1.1 Filter anlegen

An einfachsten lassen sich Filter anlegen, indem in der Metadatendarstellung auf einen Metadatenwert geklickt wird.

Über der Spalte mit den Kommunikationen wird dann der gerade erstellte Filter angezeigt, in der Tabelle werden alle Kommunikationen herausgefiltert, die nicht über dieses Attribut-Wert-Paar verfügen.

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

Filter		
[transcription-convention] [HIAT (vereinfacht)]		
<input type="button" value="Filter"/> <input type="button" value="Alle Filter löschen"/> 9 Kommunikationen		
Kommunikation		
S	Name	transcription-convention
	Helge Schneider: Arbeitsamt	HIAT (vereinfacht)
	Hubert Fichte: Interview	HIAT (vereinfacht)
	Helge Schneider: Tropfsteinhöhle	HIAT (vereinfacht)
	Studio Braun: English Translator	HIAT (vereinfacht)
	Deutsche Bahn: Rossau	HIAT (vereinfacht)
	Studio Braun: Forum Waffenrecht	HIAT (vereinfacht)
	Ségolène Royal: Débat	HIAT (vereinfacht)
	Fernando Savater: Hermandad	HIAT (vereinfacht)
	Paul McCartney: Interview	HIAT (vereinfacht)

Rechts neben einer kurzen Referenz auf den Filter befinden sich drei Schaltflächen:



Die erste Schaltfläche aktiviert- oder deaktiviert den Filter.

Die zweite Schaltfläche invertiert den Filter. Im obigen Beispiel würde ein Klick auf die Schaltfläche dazu führen, dass nur noch Kommunikationen angezeigt werden, deren „transcription-convention“ nicht „**HIAT (vereinfacht)**“ ist.

Die dritte Schaltfläche bestimmt, ob dieser Filter mit weiteren Filtern (sofern vorhanden) mit logisch UND oder mit logisch ODER verknüpft wird. Wenn ein weitere Filter beispielsweise nur Kommunikationen zulassen würde, in denen die Sprache DEU (=Deutsch) vorkommt, so würde

mit dieser Schaltfläche bestimmt, ob die Bedingungen beide erfüllt sein müssen oder nur eine von beiden.

6.1.2 Filter modifizieren

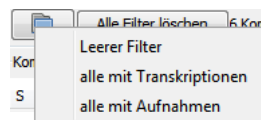
Die Filter selber bestehen aus einem XPath-Ausdruck, der auf das XML der Coma-Datei angewendet wird. Dieser XPath-Ausdruck lässt sich von fortgeschrittenen Anwendern modifizieren, in dem man in die Beschreibung des Filters hinein klickt:

[Description/Key[@Name='transcription-convention' and text()='HIAT (vereinfacht)']]

Das Feld färbt sich daraufhin gelb und Sie können den verwendeten XPath modifizieren. Mit einem Druck auf die **Enter**-Taste wird der XPath übernommen. Handelt es sich um einen ungültigen XPath, so färbt sich das Eingabefeld rot und der Filter wird nicht übernommen.

6.1.3 Vorgefertigte Filter

Für Kommunikationen gibt es einige vorgefertigte Filter. Sie lassen sich über das Ordnersymbol in der rechten Spalte auswählen:



- **Leerer Filter:** Erstellt einen Filter mit einem leeren XPath, der dann vom Anwender modifiziert werden kann.
- **alle mit Transkriptionen:** Erstellt einen Filter, der nur Kommunikationen anzeigt, für die Transkriptionen vorliegen.
- **alle mit Aufnahmen:** Erstellt einen Filter, der nur Kommunikationen anzeigt, für die Aufnahmen vorliegen.

6.1.4 Filter löschen

Die Schaltfläche „**Alle Filter löschen**“ löscht alle Filter ohne Rückfrage. Einzelne Filter lassen sich nur deaktivieren, nicht löschen.

6.1.5 Filter bei Sprechern

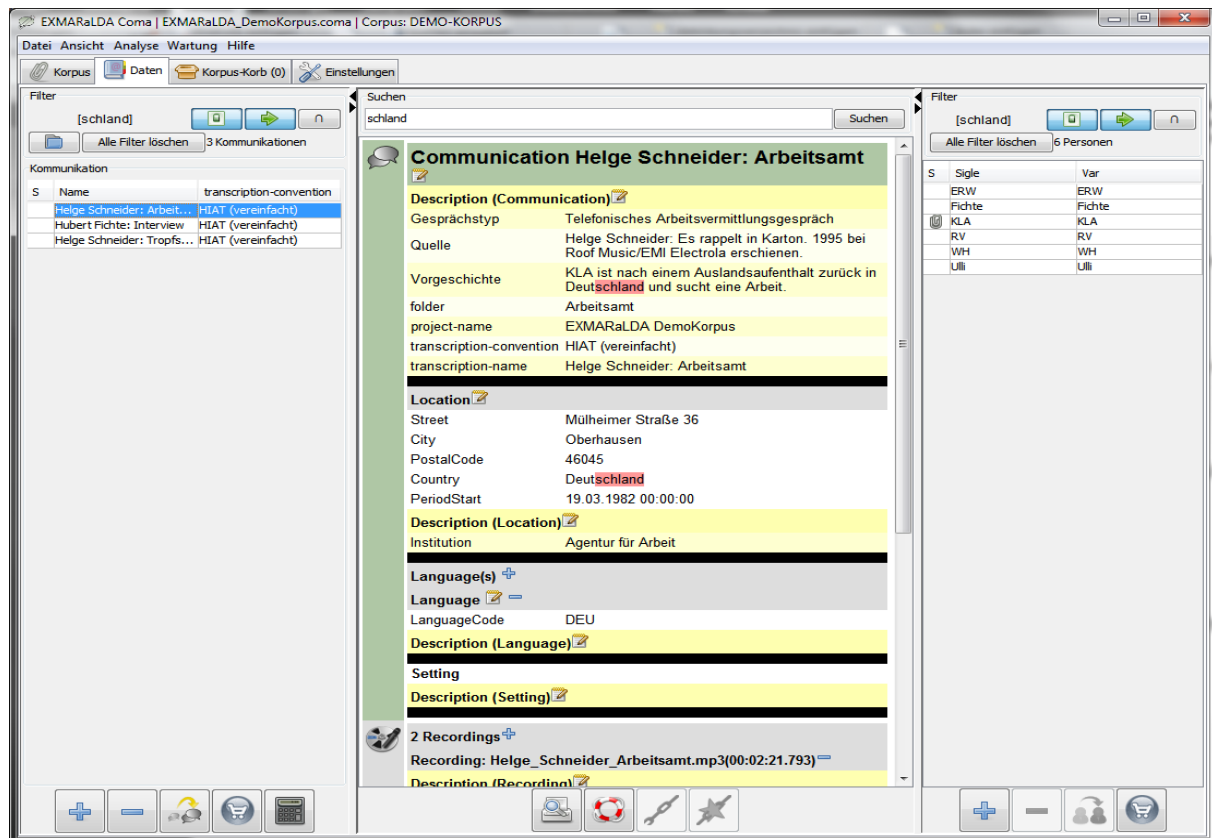
Für die Sprecher-Spalte funktionieren die Filter analog zu denen der Kommunikationen-Spalte, es gibt jedoch keine vorgefertigten Filter.

6.2 Suchen

Eine erweiterte Form des Filterns lässt sich über das Suchfeld oberhalb der Metadaten-Spalte erreichen.

Wird in das Feld eine Zeichenfolge eingegeben und auf den **Suchen**-Button geklickt, so wird sowohl bei den Kommunikationen als auch bei den Sprechern ein Filter eingerichtet, der nur die Elemente anzeigt, die diese Zeichenfolge in irgendeinem Metadatenwert enthalten.

Zusätzlich wird die gesuchte Zeichenfolge in der Metadatenanzeige farblich hervorgehoben. Im hier abgebildeten Beispiel wurde nach der Zeichenfolge „schland“ gesucht; im Screenshot ist zu sehen, dass in der markierten Kommunikation zwei Metadatenwerte die Zeichenfolge enthalten (jeweils Deutschland).



6.3 In Transkriptionen suchen

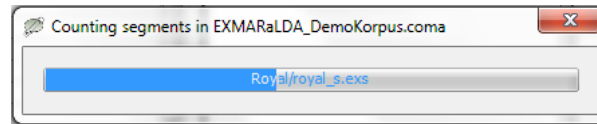
Aus Coma lassen sich Suchabfragen über alle Transkriptionen des geladenen Korpus anstoßen. Wählen Sie dazu die **Analyse > Mit EXAKT im Korpus suchen**. Weitere Informationen zur Suche in EXAKT finden Sie in der EXAKT-Anleitung, die Sie unter <http://www.exmaralda.org/>, Sektion „Hilfe/Support“ herunterladen können.

6.4 Drucken der Metadatenansicht

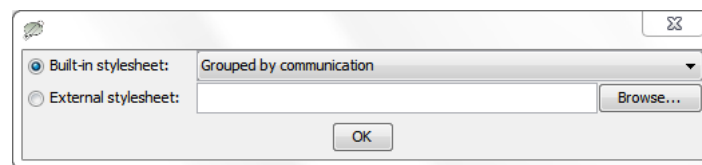
Mit einem Klick auf das Druckersymbol unter der Metadatenansicht wird die aktuell sichtbare Metadatenansicht (Kommunikation, Kommunikation+Sprecher oder Sprecher) an einen Webbrowser geschickt, von wo aus sie sich ausdrucken lässt. Beachten Sie, dass sie die Ansicht über **Ansicht > schwarz/weiß** umschalten können, um möglicherweise eine bessere Druckansicht zu erhalten.

6.5 Korpusstatistik erstellen

Über die Option **Analyse > Korpusstatistik generieren** lässt sich Statistik für die Segmentzahlen des geöffneten Korpus erstellen, und zwar entweder nach Kommunikationen oder nach Sprechern. Wenn Sie die Option aufrufen, werden zunächst die Segmentzahlen aller verknüpften Transkriptionen ermittelt:



Dann erscheint ein Dialog, in dem Sie auswählen können, ob die Segmentzahlen nach Kommunikationen oder nach Sprechern ausgegeben werden sollen. Sie können auch ein eigenes, externes Stylesheet bestimmen, das die Auswertung übernehmen soll.



Nach einem Klick auf die **OK**-Schaltfläche wird die Korpusstatistik generiert und in einem Webbrowser geöffnet:

Ays1

Comm	Contr	Utterances	Words	Non-Pho
0200	93	166	542	122
0438	422	621	2247	398
0438a	248	404	1756	321
Total	763	1191	4545	841

Azi

Comm	Contr	Utterances	Words	Non-Pho
0302a	36	50	113	14
0302a	28	42	126	45
0302a/b	63	80	337	87
0305a	13	30	106	45
0387a/b	872	1285	4139	705
0390	429	543	1257	219
0400a/b	540	1103	3079	1220
0400b	133	229	677	211
0401a	284	500	1180	524
Total	2398	3862	11014	3070

Ban

Comm	Contr	Utterances	Words	Non-Pho
0403a	64	104	359	73
Total	64	104	359	73

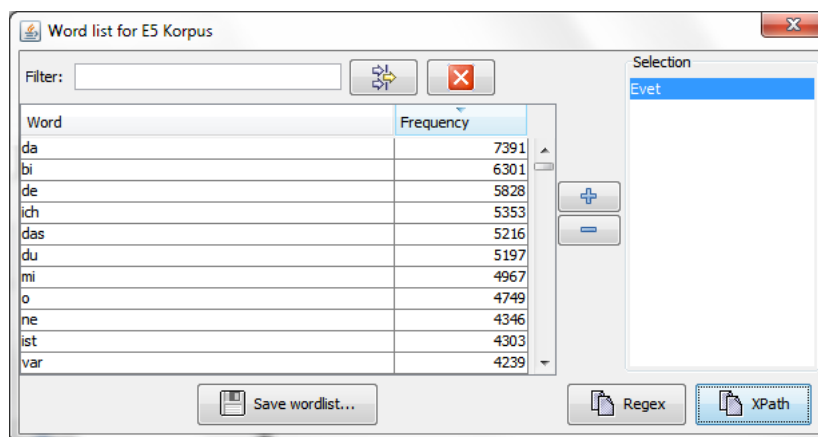
In diesem Beispiel wurde die Statistik nach Sprechern erstellt: Für alle Sprecher (hier mit Kürzeln repräsentiert) wird eine Tabelle mit allen Kommunikationen, an denen Sie teilgenommen haben, und einer genauen Aufschlüsselung der enthaltenen Segmente ausgegeben.

Wenn Sie die Auswertung in einem anderen Programm weiterverwenden wollen, so können Sie den Inhalt des Browserfensters entweder über die Zwischenablage z.B. in Word einfügen oder die HTML-Datei aus dem Browser speichern.

6.6 Wortliste erstellen

Mit der Option **Analyse > Wortliste erstellen** lässt sich eine Liste aller Wörter in allen Transkriptionen des geöffneten Korpus erstellen. Wenn Sie die Option anwählen, werden zunächst alle Wörter aller Transkriptionen (sofern sie segmentiert sind) ermittelt. Dies kann je nach Korpusgröße durchaus eine Weile dauern.


Anschließend wird die Wortliste angezeigt:



In der Tabelle auf der linken Seite sehen Sie in der linken Spalte die gefundenen Wörter und in der rechten Spalte die Anzahl der Vorkommen im Korpus. Beide Spalten lassen sich durch Klick auf den Spaltenkopf sortieren.

Mit Hilfe des „**Filter**“-Eingabefeldes am oberen Rand des Fensters lässt sich die Anzeige auf bestimmte Wörter reduzieren.

Unter der Tabelle lässt sich die Wortliste über die Schaltfläche **Save wordlist...** als HTML-Datei abspeichern, die sich u.a. auch in MS Word öffnen lässt.

In die rechte, mit „**Selection**“ überschriebene Spalte lassen sich durch Klick auf die  Schaltfläche ausgewählte Wörter in eine Auswahlliste überführen.

Mit einem Klick auf die Schaltfläche **XPath** wird ein XPath erzeugt und in die Zwischenablage kopiert, der sich als Suchausdruck im Suchwerkzeug EXAKT verwenden lässt.

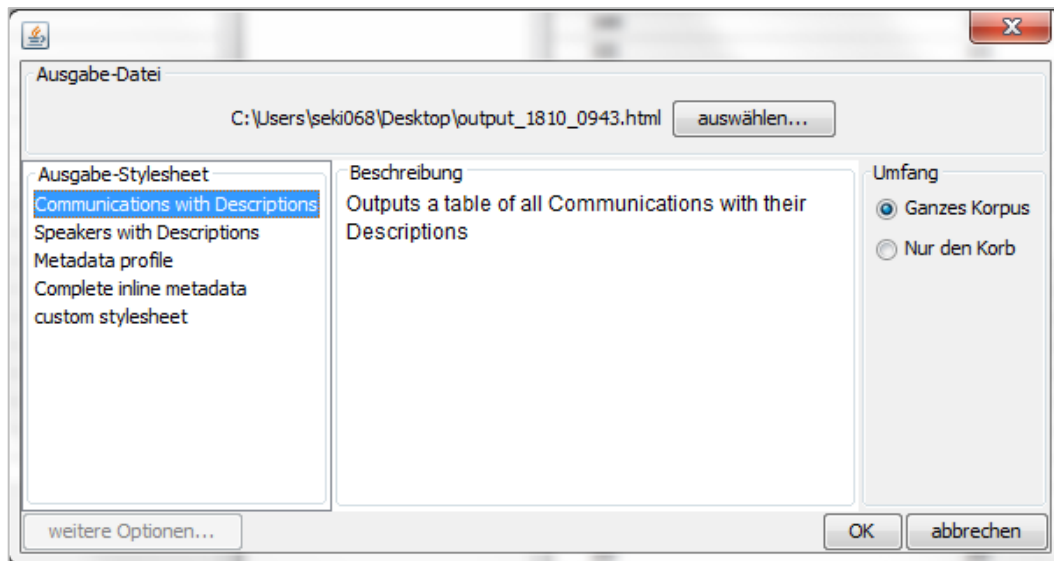
Ein Klick auf die Schaltfläche **Regex** kopiert einen regulären Ausdruck in die Zwischenablage, der sich ebenfalls in EXAKT zur Suche nach den ausgewählten Wörtern einsetzen lässt.

6.7 Wortliste für ausgewählte Sprecher erstellen

Diese Option funktioniert grundsätzlich genau wie *Wortliste erstellen*, nur müssen zuvor die Sprecher ausgewählt werden, für die die Wortliste erstellt werden soll.

6.8 Ausgabe als html-Datei

Ein Klick auf **Datei>Ausgabe...** erstellt eine html-Ausgabedatei, deren Speicherplatz selbst gewählt werden kann. Es öffnet sich das folgende Fenster:



Hier kann beispielsweise eine Tabelle aller Kommunikationen und der zugehörigen Descriptions erstellt werden. Es kann entweder ein vorgegebener Ausgabe-Stylesheet aus dem Menü oder ein eigener Stylesheet (custom stylesheet) ausgewählt werden. Außerdem kann beim Umfang ausgewählt werden, ob die Ausgabedatei das gesamte Korpus oder den Korpus-Korb miteinbeziehen soll. Bei Klick auf die vorgegebenen Stylesheets wird daneben eine Beschreibung angezeigt.

6.9 Korpus-Korb

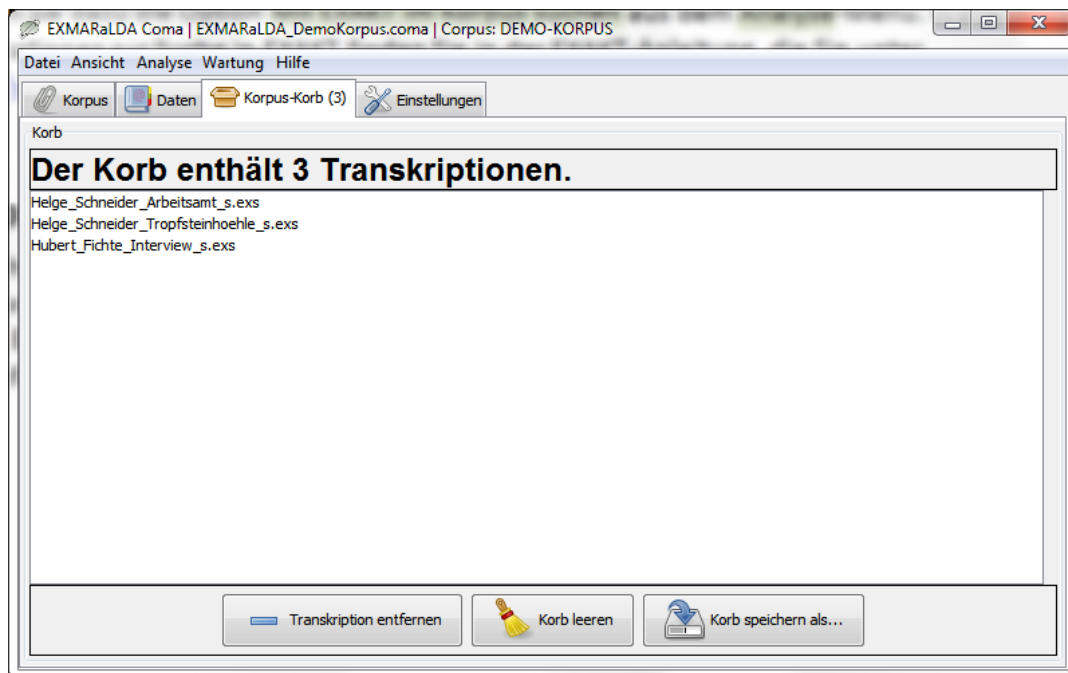
Um anhand von Metadaten eine Auswahl an Transkriptionen vornehmen zu können, die dann einer weiteren Analyse zugeführt werden können, gibt es in Coma den sogenannten „**Korpus-Korb**“. Es handelt sich dabei um eine Liste von Transkriptionen, die man – analog zu einem Einkaufskorb in einem Online-Shop – nach und nach füllen kann.

Den aktuellen Inhalt des Korpus-Korbes kann man sich anschauen, wenn man auf den entsprechenden Reiter am oberen Rand von Coma klickt.

6.9.1 Verwalten des Korbes

Der Korpus-Korb-Reiter bietet drei Schaltflächen zur Verwaltung:

- **Transkription entfernen:** Wenn in der Liste eine Transkription ausgewählt ist, so lässt sie sich mit einem Klick auf diese Schaltfläche aus dem Korb entfernen. Die Transkription wird nur aus dem Korb genommen; am eigentlichen Korpus ändert sich dadurch nichts.
- **Korb leeren:** Entfernt alle Transkriptionen aus dem Korb.



- **Korb speichern als:** Mit dieser Schaltfläche können Sie den Inhalt des Korbes auf dem Computer abspeichern. Der Korb wird dabei als eine vollwertiges Coma-Korpus gespeichert und lässt sich anschließend sowohl in Coma als auch in EXAKT öffnen, enthält aber eben nur die Metadaten zu den Transkriptionen, die vorher in den Korb gelegt wurden.

Um die volle Funktionsfähigkeit des gespeicherten Korpus zu erhalten, sollten Sie es in dasselbe Verzeichnis speichern wie die Coma-Datei, aus dem Sie es generiert haben.

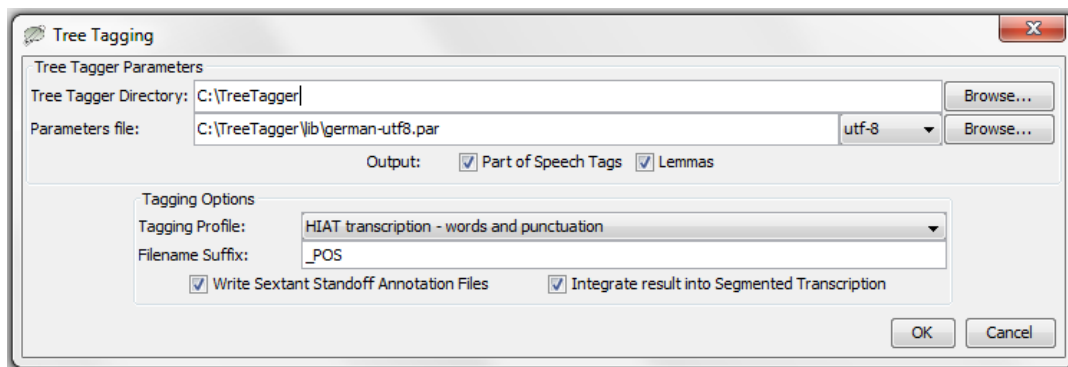
6.10 TreeTagger

In Coma gibt es zusätzlich die Möglichkeit, die eingefügten segmentierten Transkriptionen direkt mit dem TreeTagger (siehe <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>) zu taggen. Hierzu muss zunächst das Programm TreeTagger über den angegebenen Link installiert werden. Bitte beachten Sie die Hinweise zur Installation auf den jeweiligen Betriebssystemen. Bei der Installation für Windows ist es wichtig, dass der TreeTagger-Ordner direkt beim Laufwerk C:\ eingefügt wird. Der entsprechende Pfad sollte dann C:\TreeTagger heißen.

Die Parameter-Dateien für die entsprechenden Sprachen, die mit dem TreeTagger annotiert werden sollen, müssen zusätzlich von der Webseite herunter geladen werden. Anschließend müssen sie in den richtigen Ordner der TreeTagger-Datei eingefügt werden, wie es auf der Webseite erklärt wird.

Wenn die Installation vollständig ist, kann der TreeTagger direkt aus der Coma-Datei über **Werkzeuge>TreeTagger...** aufgerufen werden. Danach muss als Directory der Pfad des TreeTagger-Ordners ausgewählt werden. Als Parameters file muss die Parameter-Datei der Sprache, die im Korpus annotiert werden soll, ausgewählt werden. Hinter der Parameter-Datei muss die Kodierung des Parameters festgelegt werden. Bei den Parametern auf der Webseite handelt es sich vor allem um UTF-8-kodierte Parameter.

Beim „**Output:**“ kann ausgewählt werden, ob entweder nur „**Part of Speech Tags**“ oder „**Lemmas**“ oder beides ausgegeben werden soll.



Bei den „**Tagging Options**“ kann ein „**Tagging Profile**“, das sich auf die verwendete Transkriptionskonvention bezieht, ausgewählt werden. Außerdem ist das Suffix, das die neuen Dateien erhalten sollen, frei wählbar.

Durch die anwählbaren Kästchen unterhalb des Suffixes lässt sich das Ergebnis des TreeTaggers modifizieren. Es kann ausgewählt werden, ob eine Standoff-Annotation-Datei (.ESA-Datei) mit den Ergebnissen, die sich mit Sextant öffnen lässt, erstellt werden soll. Außerdem kann entschieden werden, ob die Ergebnisse in die segmentierte Transkription (.exs-Datei) eingefügt werden soll.

Wenn das Tagging mit dem TreeTagger erfolgreich war, erhält man also je nach Auswahl eine neue .exs- und .esa-Datei mit dem ausgewählten Suffix, die die Ergebnisse des TreeTaggers enthalten. Diese Dateitypen lassen sich unter anderem mit dem EXMARaLDA Programm Sextant (<http://exmaralda.org/de/sextant/>) öffnen.

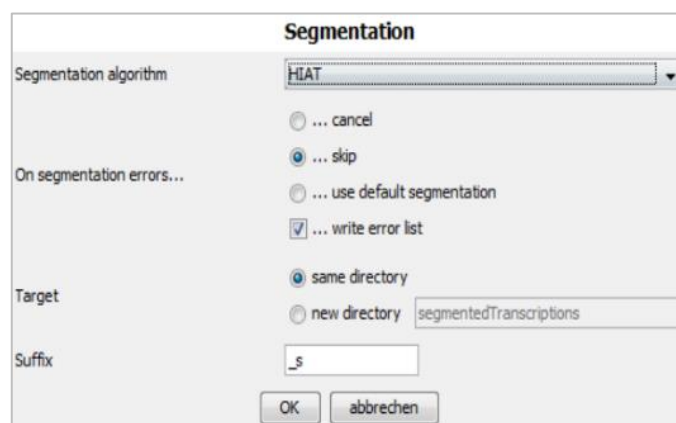
7 Wartung

Coma bietet einige Optionen zur Datenpflege sowohl der mit dem Korpus verknüpften Transkriptionen als auch der Metadaten.

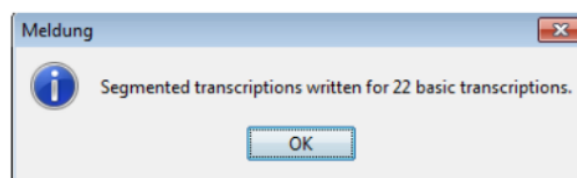
7.1 Überprüfung verknüpfter Transkriptionen

7.1.1 Transkriptionen segmentieren

Im Allgemeinen finden alle Änderungen an Transkriptionen im Partitur-Editor auf der Grundlage von den Basistranskriptionen (Dateiendung .exb) statt. Das segmentierte Transkriptionsformat („.exs“) dient hingegen zum Integrieren der Transkriptionen in einen Korpus in Coma (.coma Datei) und als Grundlage für das Suchen in EXAKT. Wann immer sie Fehlerkorrekturen und Änderungen in Ihrer .exb Datei speichern, können Sie Ihre segmentierte(n) Transkription(en) im Partitur-Editor (über: **Transkription > Segmentierte Transkription exportieren...**) aktualisieren. Sie können Fehler auch zuerst in Ihren .exb Dateien korrigieren und dann die folgenden Coma Funktionen verwenden. Hierfür wählen Sie „**Transkriptionen segmentieren**“ im Menüpunkt Wartung an und setzen Ihre Einstellungen (d.h. Segmentierungsalgorithmus, Zielverzeichnis, Suffix2 etc.).



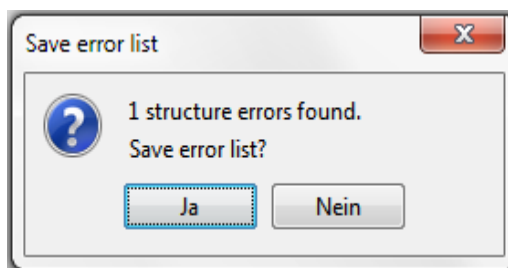
Bestätigen Sie die Eingaben mit OK. Die .exs Dateien für alle Basistranskriptionen in ihrem Korpus werden automatisch generiert. Der folgende Dialog wird angezeigt.



Klicken sie zum Bestätigen **OK**.

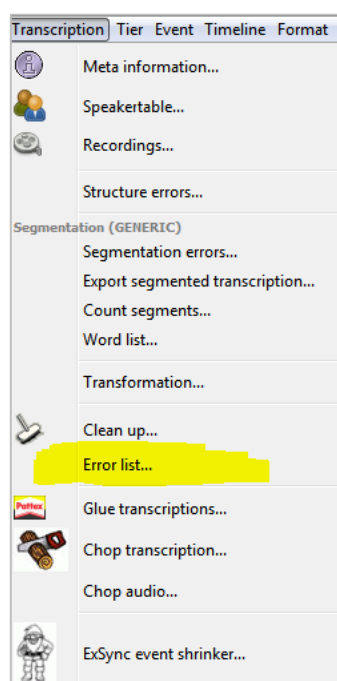
7.1.2 Auf Strukturfehler überprüfen

Diese Funktion überprüft alle verknüpften Transkriptionen des Korpus auf eventuelle Strukturfehler. Wird in einer oder mehreren Transkriptionen ein solcher Fehler gefunden, so erscheint eine Meldung, aus der heraus man eine Fehlerliste für das Korpus speichern kann:

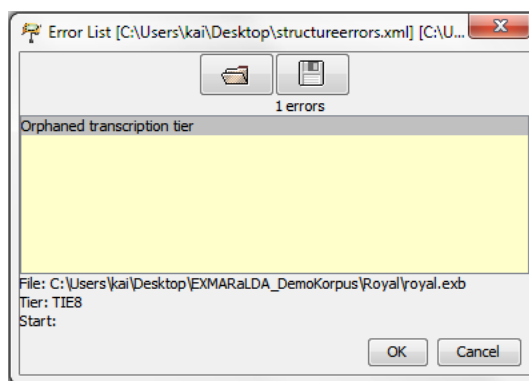


7.1.3 Fehlerliste

Bei dieser Fehlerliste handelt es sich um eine XML-Datei, die sich – nachdem man sie auf seinem Computer gespeichert hat – im Partitureditor öffnen lässt. Die entsprechende Option befindet sich im **Transcription**-Menü des Partitureditors:



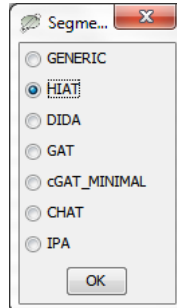
Im sich öffnenden Fenster kann man durch Druck auf das „Datei öffnen“-Symbol die eben gespeicherte Fehlerliste öffnen und bekommt dann eine Liste der gefundenen Fehler.



Ein Doppelklick auf einen Fehler (hier eine verwaiste Transkriptionsspur) öffnet die entsprechende Transkription und springt an die Stelle mit dem Fehler.

7.1.4 Auf Segmentierungsfehler überprüfen

Analog zur Suche nach Strukturfehlern lässt sich auch nach Segmentierungsfehlern suchen. Wird die Funktion aufgerufen, muss zunächst bestimmt werden, nach welchem Segmentierungsalgorithmus die Transkriptionen segmentiert wurden:

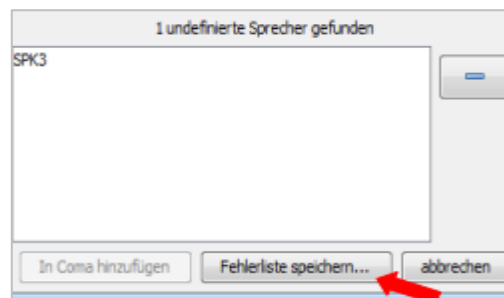


Das weitere Vorgehen entspricht dem bei der Suche nach Strukturfehlern.

7.1.5 Sprecherzuordnung prüfen

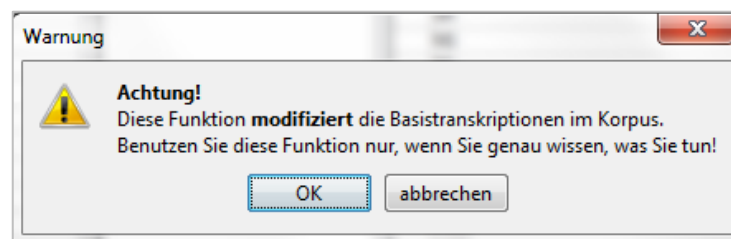
Diese Option sucht nach „undefinierten Sprechern“ (d.h. Sprecher, denen keine Spur zugeordnet ist). Es öffnet sich ein neuer Dialog mit den Ergebnissen.

Die Fehlerliste kann genau wie bei den Strukturfehlern gespeichert werden.

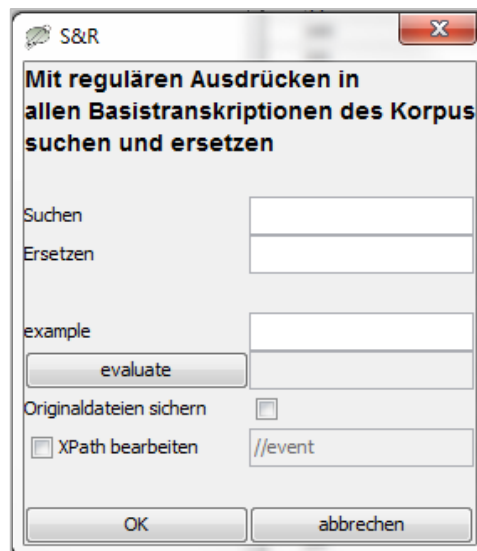


7.1.6 Suchen und Ersetzen im Korpus

Diese Funktion sucht und ersetzt reguläre Ausdrücke in den ins Korpus eingefügten Basistranskriptionen. Nach der Auswahl der Funktion erscheint eine Warnmeldung:



Klicken Sie zum Fortfahren **OK**, dadurch erscheint folgender Dialog:



Jetzt können Sie Elemente in allen Basistranskriptionen des Korpus suchen und ersetzen. Beachten Sie, dass Sie das Feld Originaldateien sichern anwählen können, falls Sie die Ursprungsdatei speichern möchten. Bestätigen Sie mit **OK**. Am Ende erhalten Sie eine Zusammenfassung aller vorgenommenen Änderungen. Falls nötig, kopieren Sie die Ergebnisse und fügen Sie diese in einen Editor oder Word ein.

7.2 Pflege der Metadaten

7.2.1 Segmentzahlen aktualisieren

Extrahiert aus allen mit dem Korpus verknüpften (segmentierten) Transkriptionen die Segmentzahlen und schreibt sie als geschützte Metadaten in deren Description.

Beim Verknüpfen von Transkriptionen werden die Segmentzahlen ebenfalls in die Description geschrieben; die Funktion aus dem **Wartung**-Menü ist nützlich, wenn sich Transkriptionen (und Segmentzahlen) geändert haben.

7.2.2 Aufnahmen aus Transkriptionen aktualisieren

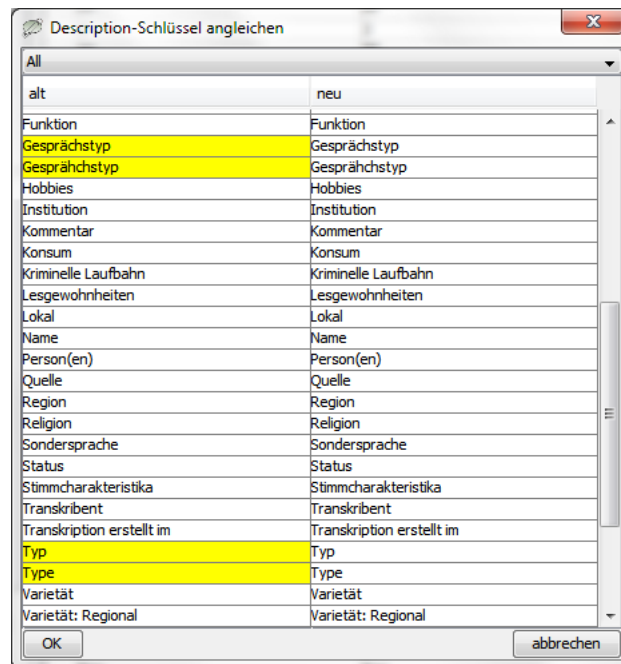
Diese Funktion schaut in den Kopfdaten aller Transkriptionen im Korpus nach zugewiesenen Mediendateien und fügt diese als Recording in die Coma-Datei ein. In Coma bereits existierende Recordings werden nicht verändert oder gelöscht.

Diese Funktion ist insbesondere nützlich, wenn bereits verknüpften Transkriptionen nachträglich weitere Mediendateien (z.B. in anderen Audio- oder Videoformaten) zugewiesen werden.

7.2.3 Description-Schlüssel angleichen

Häufig wird die Vergabe von Schlüsselnamen nicht einheitlich geregelt oder bei der Eingabe passieren Flüchtigkeitsfehler – dies führt dazu, dass die entsprechende Information später nicht aufgefunden werden kann. Die Funktion „**Description-Schlüssel angleichen**“ dient dazu, die Schlüsselnamen, die in den Descriptions in Coma verwendet werden, zu vereinheitlichen.

Wird die Funktion aufgerufen, erscheint eine Tabelle mit allen Schlüsselnamen, die in Descriptions des geöffneten Korpus verwendet wurden.

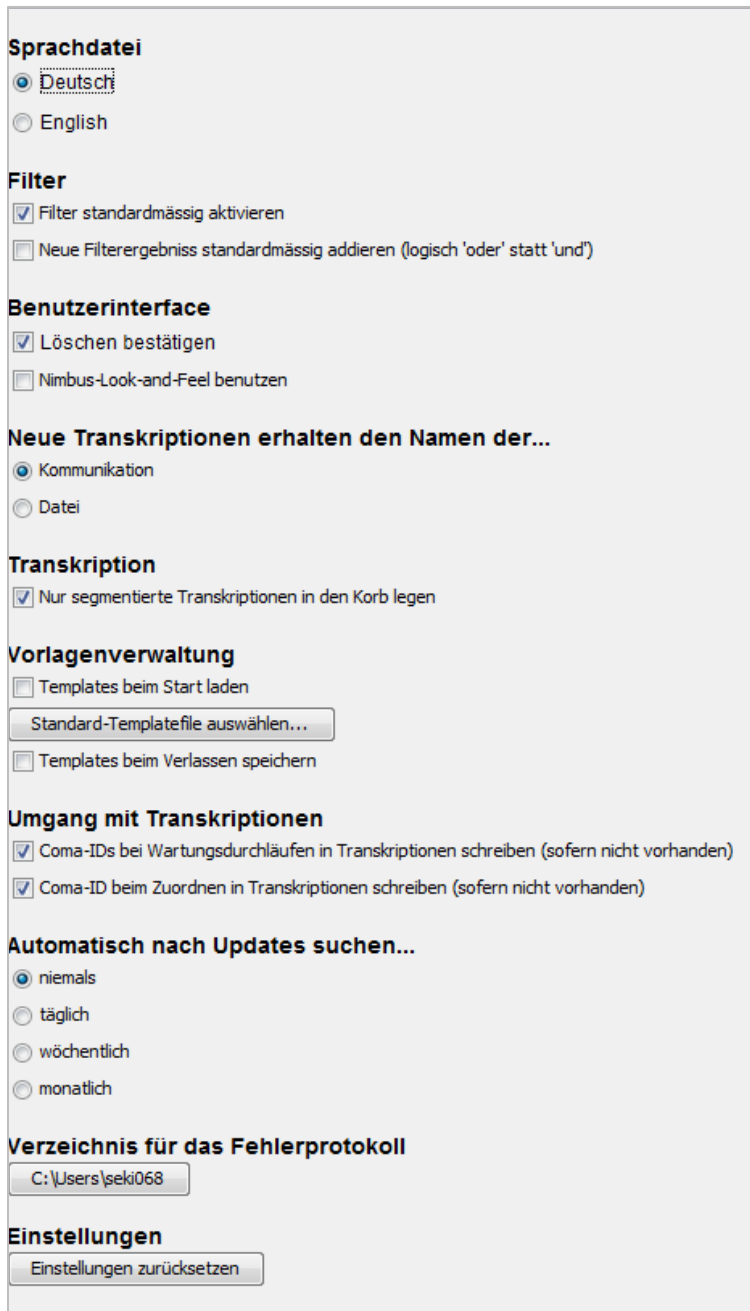


Schlüsselnamen, die sich ähneln (und damit einen Hinweis auf einen möglichen Fehler geben), werden mit gelber Farbe hervorgehoben.

In der rechten Spalte lassen sich die Schlüsselnamen editieren. Wird also im obigen Beispiel in der rechten Spalte „**Gesprächstyp**“ durch „**Gesprächstyp**“ ersetzt, so wird (nach drücken der **OK**-Schaltfläche) der entsprechende Schlüsselname in allen Descriptions des Korpus ersetzt.

8 Einstellungen

Über das Menü **Bearbeiten>Einstellungen** können Sie die Voreinstellungen des Programms ändern. Die einstellbaren Optionen sind im Folgenden erläutert:



Sprachdatei

☒ Deutsch

☐ English

Filter

☒ Filter standardmässig aktivieren

☐ Neue Filterergebniss standardmässig addieren (logisch 'oder' statt 'und')

Benutzerinterface

☒ Löschen bestätigen

☐ Nimbus-Look-and-Feel benutzen

Neue Transkriptionen erhalten den Namen der...

☒ Kommunikation

☐ Datei

Transkription

☒ Nur segmentierte Transkriptionen in den Korb legen

Vorlagenverwaltung

☐ Templates beim Start laden

Standard-Templatefile auswählen...

☐ Templates beim Verlassen speichern

Umgang mit Transkriptionen

☒ Coma-IDs bei Wartungsdurchläufen in Transkriptionen schreiben (sofern nicht vorhanden)

☒ Coma-ID beim Zuordnen in Transkriptionen schreiben (sofern nicht vorhanden)

Automatisch nach Updates suchen...

☒ niemals

☐ täglich

☐ wöchentlich

☐ monatlich

Verzeichnis für das Fehlerprotokoll

C:\Users\seki068

Einstellungen

Einstellungen zurücksetzen

- **Sprachdatei:** An dieser Stelle lässt sich die Sprache der Benutzeroberfläche von Coma einstellen.

ACHTUNG: Die Bezeichnungen der verschiedenen Oberflächenelemente werden in einer Ressourcendatei verwaltet - die Deutsche und Englische Version ist weitestgehend vollständig, für die anderen angebotenen Sprachen sind wir auf die Hilfe der Anwender angewiesen. Wenn sie also gerne bei der Lokalisierung von Coma helfen wollen, so melden Sie sich bei Kai Wörner [kai.woerner@uni-hamburg.de].

Wenn Sie die Sprache der Benutzeroberfläche geändert haben, müssen Sie das Programm neustarten, um die Änderungen zu sehen.

- **Filter:** An dieser Stelle lässt sich einstellen, wie *Filter* für die Kommunikationen- und Sprechertabellen angelegt werden:
 - **„Filter standardmässig aktivieren“:** Mit dieser Einstellung werden Filter sofort angewendet, wenn sie hinzugefügt werden.
 - **„Neue Filterergebnisse standardmässig addieren (logisch ‚oder‘ statt ‚und‘)“:** bestimmt, ob beim Hinzufügen von weiteren Filtern die Ergebnisse logisch ‚oder‘ bzw. ‚und‘ verknüpft werden. Nähere Erläuterungen zu den Auswirkungen finden Sie im Kapitel *Filter*.
- **Benutzerinterface**
 - **„Löschen bestätigen“:** Hier können Sie festlegen, ob eine Sicherheitsabfrage erfolgt, wenn Sie Elemente aus den Metadaten löschen wollen.
- **Neue Transkriptionen erhalten den Namen der...:** Wenn einer Kommunikation eine Transkription zugeordnet wird (siehe *Verwaltung von Metadaten*), so erhält diese einen Namen; an dieser Stelle lässt sich auswählen, ob dieser Name dem Dateinamen oder dem Namen der Kommunikation entsprechen soll. Der Dateiname der Transkription wird davon nicht beeinflusst. Wird die Einstellung „Kommunikation“ gewählt und eine Kommunikation hat mehrere zugeordnete Transkriptionen, so werden die Namen aufsteigend durchnummeriert.
- **Transkription** Im EXMARaLDA-System lassen sich die meisten Operationen (Suchen, Visualisierungen etc.) nur mit segmentierten Transkriptionen durchführen. An dieser Stelle können Sie festlegen, ob Basistranskriptionen in Coma überhaupt berücksichtigt werden:
 - **„Nur segmentierte Transkriptionen in den Korb legen“:** Wenn sie diese Option einschalten, so werden nur segmentierte Transkriptionen in den Korpus-Korb gelegt, unabhängig davon, ob Kommunikationen auch Basistranskriptionen zugeordnet sind.
- **Vorlagenverwaltung** Für bestimmte Datentypen in Coma lassen sich *Templates* (Vorlagen) anlegen. An dieser Stelle in den Einstellungen können Sie bestimmen, ob eine Template-Datei beim Start von Coma automatisch geladen werden soll und beim Verlassen automatisch wieder gespeichert wird. Mit der Schaltfläche zwischen diesen Optionen können Sie die entsprechende Template-Datei auswählen. Die Option zum automatischen Laden einer Template-Datei findet sich übrigens auch im **Datei**-Menü des Programms.

9 Anmerkungen und Haftung

9.1 Haftungsbegrenzung

In keinem Fall, außer wenn durch geltendes Recht gefordert oder schriftlich zugesichert, ist irgendein Urheberrechtsinhaber oder irgendein Dritter, der das Programm wie oben erlaubt modifiziert oder übertragen hat, Ihnen gegenüber für irgendwelche Schäden haftbar, einschließlich jeglicher allgemeiner oder spezieller Schäden, Schäden durch Seiteneffekte (Nebenwirkungen) oder Folgeschäden, die aus der Benutzung des Programms oder der Unbenutzbarkeit des Programms folgen (einschließlich – aber nicht beschränkt auf – Datenverluste, fehlerhafte Verarbeitung von Daten, Verluste, die von Ihnen oder anderen getragen werden müssen, oder dem Unvermögen des Programms, mit irgendeinem anderen Programm zusammenzuarbeiten), selbst wenn ein Urheberrechtsinhaber oder Dritter über die Möglichkeit solcher Schäden unterrichtet worden war.

9.2 Preview-Versionen

Von EXMARaLDA-Versionen werden in regelmäßigen Abständen Preview-Versionen veröffentlicht (<http://exmaralda.org/de/preview-version/>), die den neuesten Entwicklungsstand der Software repräsentieren. Obwohl wir auch für die regulär veröffentlichten Versionen der Software keine Funktionsgarantie und Gewährleistung geben (können) – siehe oben – möchten wir doch darauf hinweisen, dass regulär veröffentlichte Versionen immerhin den Anspruch erheben, vor der Veröffentlichung sorgfältig getestet worden zu sein. Fehlfunktionen und unerwartetes Verhalten der Software sind bei Preview-Versionen durchaus zu erwarten und sie sollten immer eine Sicherheitskopie ihrer Daten machen, bevor Sie sie mit Preview-Software bearbeiten.

9.3 Dieses Dokument

Dieses Dokument wurde erstellt und wird gepflegt von Kai Wörner (kai.woerner@uni-hamburg.de) – es erhebt keinen Anspruch auf Vollständigkeit. Falls Ihnen Fehler auffallen oder falls Sie bestimmte Funktionen des Programms nicht erläutert sehen, so lassen Sie es mich wissen

