|  |  |  |
| --- | --- | --- |
|  |  |  |
| partitureditor2 | coma2 | exakt |
|  |  |  |
| **How to:**  **Importieren einer Texttranskription** | | |
|  | | |

Dieses Dokument erklärt, wie Sie Transkriptionen gesprochener Sprache, die mithilfe eines Texteditors oder Textverarbeitungsprogramms erstellt wurden, in den Partitur-Editor (als Simple EXMARaLDA" Format) importieren können. Simple EXMARaLDA ist ein Format für einfache Textdateien, das auch einige grundlegende Annotationen, nicht-verbale Phänomene und Überlappungen verarbeiten kann. Wie Sie einfache Textdateien oder einen TreeTagger Output in EXMARaLDA importieren, können Sie in dem Dokument „How to use the PE with written data“ nachlesen.

Bevor Sie dieses Dokument lesen, sollten Sie sich mit „Understanding the basics of EXMARaLDA“ auseinander setzen.

Inhalt

A. Vorbereitung der Datei für den Import 2

1. Struktur und Informationen der Quelldatei 2

2. Das Simple EXMARaLDA Format 3

3. Konvertierung von Dateien in das Simple EXMARaLDA Format 4

4. „Nur-Text“ (Plain text) 4

5. Hinzufügen von Spuren 4

B.Importieren der Datei in den Partitur-Editor 5

1. Nachbearbeitung (Post-Editing) 6

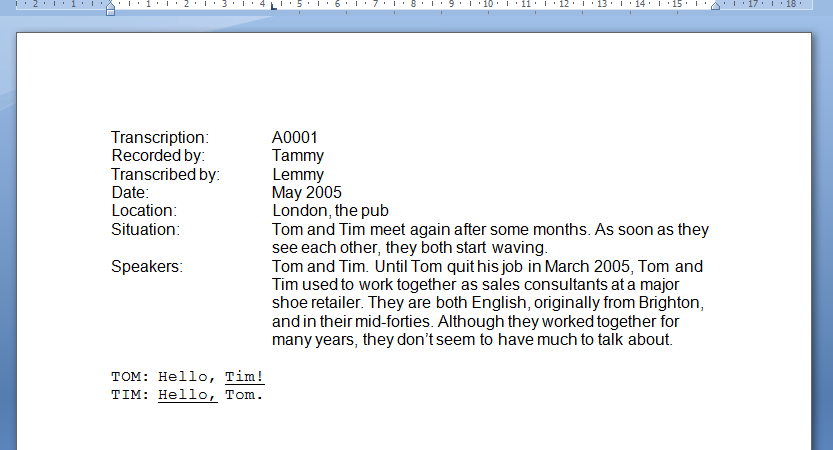
2. Metadaten 7

# A. Vorbereitung der Datei für den Import

## Struktur und Informationen der Quelldatei

Mit dem Simple EXMARaLDA Format ist es möglich, eine einfache .txt Datei so zu gestalten, dass sie vom EXMARaLDA Partitur-Editor in eine Transkription von mehreren Sprechern umgewandelt werden kann. Hierbei können Überlappungen der Sprecherspuren dargestellt und eine Annotations- und eine Deskriptionsspur definiert werden. Das Simple EXMARaLDA Format ist im Normalfall die beste Lösung, um verschiedene Altdaten zu konvertieren. Je nach Gestaltung der Transkription und den darin verwendeten Konventionen, ist die Konvertierung in dieses Format eine mehr oder weniger aufwendige Aufgabe.

Zunächst muss festgestellt werden, welches Markup verwendet wurde, um verschiedene Informationen der transkribierten Kommunikation zu beschreiben. Eine vollautomatische Konvertierung des Transkriptionsformats in das Simple EXMARaLDA Format ist nur dann möglich, wenn das Layout und/oder Markup auf eine konsistente Weise verwendet wurde, sodass unterschiedliche, verschieden kodierte Informationen ohne menschliche Interpretation entschlüsselt werden können. Mehrdeutige Annotationen oder Markups müssen manuell angepasst werden müssen.

In dem folgenden kurzen Transkript ist das einzige verwendete Markup die Unterstreichung, die zur Markierung überlappender Rede genutzt wird. Vor dem eigentlichen Beginn der Transkription gibt es im selben Dokument Metadaten über das kommunikative Ereignis und die teilnehmenden Sprecher.

Das Simple EXMARaLDA Format kann nur die Transkription verarbeiten. Speichern Sie deshalb eine Kopie der Transkription mit den entsprechenden Metadaten, damit diese nach der Überführung in den Partitur-Editor nachgetragen werden können. Da mit den EXMARaLDA-Tools Metadaten (wie die in der Kommunikation verwendeten Sprachen, Orte, Aufnahmedatum oder die L1 und L2 der einzelnen Sprecher) durchsucht und gefiltert werden können, lohnt es sich hier gewissenhaft vorzugehen.

## Das Simple EXMARaLDA Format

Damit die Textdatei vom Partitur-Editor in eine komplexe Transkription überführt werden kann, müssen die im folgenden beschriebenen Konventionen des Simple EXMARaLDA Formats eingehalten werden.

Jede Zeile beginnt mit einem individuellen *Sprecherkürzel*, gefolgt von einem *Doppelpunkt und einer Leertaste*. Dabei ist zu beachten, dass zwischen *Groß-/Kleinschreibung* unterschieden wird: „Tom“ und „TOM“ werden als zwei unterschiedliche Sprecher behandelt. In dieser Beispiel-Transkription gibt es zwei Sprecher:

TOM:

TIM:

Da jede Zeile genau einem separaten Ereignis in der EXMARaLDA-Datei entspricht, ist es nützlich, jede Äußerung in eine separate Zeile zu setzen. Jede Zeile muss mit einem Zeilenumbruch enden, zusätzliche leere Zeilen, mit zum Beispiel mehr als einem Zeilenumbruch, sind erlaubt.

TOM: Hello, Tim!

TIM: Hello, Tom.

Der Text in eckigen Klammern vor dem Text wird als paralleles Ereignis (mit korrespondierenden Start- und Endpunkten) in die **Deskriptionsspur**eingefügt. Dieses Vorgehen eignet sich für nonverbales Verhalten. In unserem Beispiel zur Beschreibung, dass beide Sprecher winken, während sie einander grüßen.

TOM: [waving] Hello, Tim!

TIM: [waving] Hello, Tom.

Der Text in geschweiften Klammern nach dem Textwird als paralleles Ereignis (mit korrespondierenden Start- und Endpunkten) in die **Annotationsspur**eingefügt. Dies ist für sprachliche Phänomene oder zusätzliche Informationen, wie z.B. Übersetzungen, geeignet. Bitte beachten Sie, dass es nur möglich ist den Text in einer Zeile als Ganzes zu annotieren, sodass in diesem Beispiel das Winken von Anfang bis Ende der Äußerungen stattfindet und keine Wort-für-Wort Übersetzung vorliegt, obwohl die Wörter in diesem konkreten Fall übereinstimmen.

TOM: [waving] Hello, Tim! {Salut, Tim!}

TIM: [waving] Hello, Tom. {Salut, Tom!}

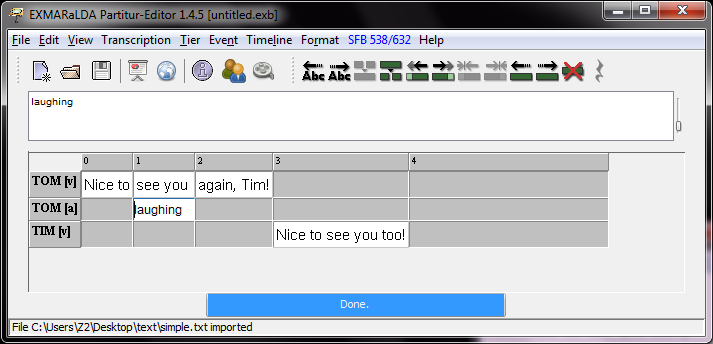
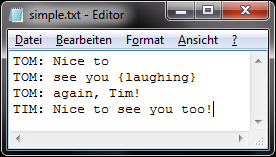
Überlappende Rede ist durch spitze Klammern markiert. Die Kennziffer (vorzugsweise eine Zahl) zwischen den beiden schließenden Klammern muss für jede Überlappung einzigartig sein und kennzeichnet so die gemeinsam gesprochenen Abschnitte.

TOM: [waving] Hello, <Tim!>1> {Salut, Tim!}

TIM: [waving] <Hello,>1> Tom. {Salut, Tom!}

Da die eckigen, geschweiften und spitzen Klammern im Simple EXMARaLDA Format diese Bedeutung haben, können sie in der Transkription nicht anderweitig verwendet werden.

## Sinnvolles Annotieren

Da sich die Annotation in den geschweiften Klammern immer auf den gesamten Text in der gleichen Zeile bezieht, sollten Sie in Erwägung ziehen die Transkription anhand von existierenden Annotationen zu splitten, um umfangreiche Nachbearbeitung zu vermeiden. In unserem Beispiel wurde Toms Äußerung auf drei Zeilen verteilt, um eine Annotation für die zwei Wörter "See you" zu erstellen.

Beachten Sie, dass nur der Text in den geschweiften Klammern am Ende der Zeile eine Annotationsspur ergibt, wobei der Text in eckigen Klammern am Zeilenanfang in eine Kommentarspur mit Typen-Bezeichnung umgewandelt wird. Das Format kann auch auf andere Weise als vordefiniert verwendet werden, dann müssen jedoch nach dem Import zunächst die Spurtypen (via Tier Properties im Menü Tier) geändert werden. Erst dann ist es möglich die Werkzeuge wie gewohnt zu nutzen.

## Konvertierung von Dateien in das Simple EXMARaLDA Format

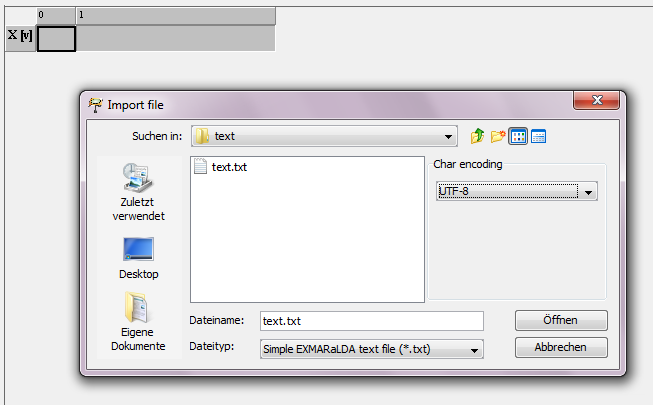
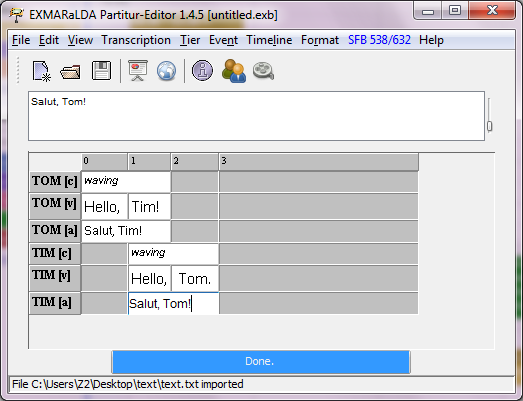
Da die Konvertierung in das Simple EXMARaLDA Format von dem ursprünglichen Dateiformat und den Transkriptionskonventionen abhängt, kann die Transformation von Transkriptionsdateien in das Simple EXMARaLDA Format nicht allgemeingültig beschrieben werden. Obwohl die Konvertierung vorzugsweise automatisch erfolgen sollte, ist jede solche automatische Konvertierung von Transkriptionen immer ein wenig riskant. Selbst wenn die Konvertierungsschritte fehlerfrei im Hinblick auf die Transkriptionskonventionen sind, können andere Fehler oder Inkonsistenzen übersehen worden sein. Dies kann zur Folge haben, dass es Fehler in der Transkription gibt, die wiederum die Inhalte der konvertierten Datei ändern. Für komplexe Transkriptionsformate von unbekannter Qualität kann der Aufwand der notwendigen Nachbearbeitung und für das Definieren vom Konvertierungsprozess eines komplexen Dateiformates zu hoch sein. In diesen Fällen ist es vorteilhafter, sich nur auf bestimmte Teile der Formate zu konzentrieren und z.B. die meisten Annotationen manuell einzutragen.

Für den Import in den Partitur-Editor sind Dateien im „Nur-Text“ ("Plain Text", Suffix .txt) Format und nicht etwa Word, PDF o.Ä. erforderlich. Formatierungen wie Farbe oder Unterstreichung werde nicht im .txt-Format gespeichert, darum müssen diese noch vor der Konvertierung mit dem entsprechenden Simple EXMARaLDA Markup ersetzt werden. Microsoft Word und OpenOffice verfügen über eine Option zu regulären Ausdrücken (in der Suchen und Ersetzen Funktion), die Suchen und Ersetzen von Formatierung sowie Verwenden von gefundenen Ausdrücken als Teil des zu ersetzenden Ausdrucks (z.B. hinzufügen von Start-und End-Tags) ermöglicht.

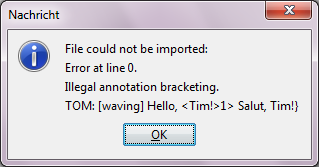
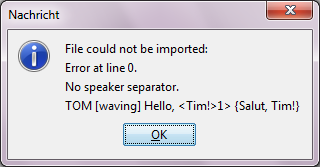
# B. Arbeit im Partitur-Editor

## Importieren der Datei

Der Import des Textes in den Partitur-Editor erfolgt über den Punkt Import aus dem Datei Menü. Suchen Sie zunächst die zu importierende Datei und stellen Sie sicher, dass Sie den richtigen Filter ausgewählt haben, also beispielsweise für den Dateityp Simple EXMARaLDA file (\*.txt) und die entsprechende Char encoding. Letzteres sollte die gleiche Kodierung sein, die auch in der Textdatei verwendet wurde.

Wenn die ausgewählte Zeichenkodierung nicht mit der Ihrer Datei übereinstimmt, können die Sonderzeichen nach dem Import nicht mehr richtig angezeigt werden. Sollte dies geschehen, versuchen Sie Ihre Textdatei mit einer anderen Zeichenkodierung zu speichern, z.B. UTF-8. Die Zeichenkodierung können Sie beim Abspeichern Ihres Dokuments in Ihrem Texteditor unter Verwendung von Speichern unter... festlegen. Versuchen Sie dann erneut Ihre Datei mit der gewählten Zeichenkodierung zu importieren.

Sollte beim Import noch eine Formatierungskonvention verletzt sein erscheint eine Fehlermeldung. Die erste Zeile gibt an, in welcher Zeilennummer der erste Fehler auftritt. Die zweite gibt Informationen über den Fehlertyp: „no speaker separator“ wenn ein Doppelpunkt zwischen dem Sprecherkürzel und dem Text fehlt, oder „illegal annotation bracketing“ wenn eine Klammer vergessen wurde. Die dritte Zeile zeigt die fehlerhafte Zeile selbst. Achten Sie darauf, dass die Datei dem Simple EXMARaLDA Format und den oben beschriebenen Konventionen entspricht und versuchen Sie dann erneut die Datei zu importieren.

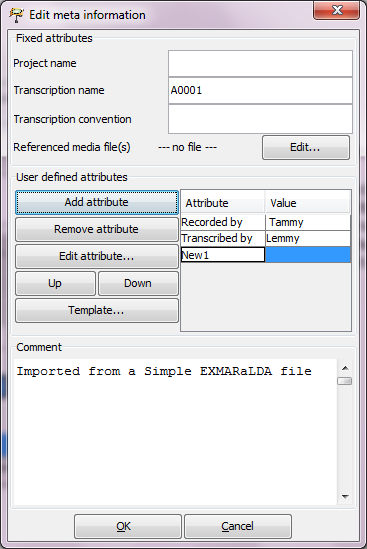
Im nächsten Schritt speichern Sie die Transkription im .exb Format (EXMARaDLA Basis-Transkription) bevor Sie mit der Eingabe von Metadaten oder dem Editieren der Transkription starten.

## Nachbearbeitung (Post-Editing)

Wenn Sie Annotations- und Kommentarspuren für Annotationen verwendet haben, müssen Sie den Spur-Typ von (D)escription in (A)nnotation ändern. Wenn Sie verschiedene Arten von Informationen in eine Annotationsspur gesetzt haben und einige davon in eine andere Spur bewegen möchten, weil Sie zum Beispiel eine Spur für Kommentare bezüglich der Aussprache wünschen, können Sie die Option Ereignisse kopieren aus (das Kontrollkästchen Copy text sollte angewählt sein) zum Hinzufügen weiterer Annotationspuren nutzen. Durch dieses Vorgehen werden die Ereignisgrenzen mit ihren Inhalten aus der ersten Spur in die nächste übertragen.

## Metadaten

Bitte vergessen Sie nicht alle Metadaten zu der Kommunikation (Transkription > Meta- Information…) und den Sprechern hinzuzufügen (Transkription > Sprechertabelle…)!



Metadaten aus der ursprünglichen Transkription werden als Attribut-Wert-Paare zur EXMARaLDA Transkription hinzugefügt.

Metadaten zu den Sprechern werden separat in die Sprechertabelle eingetragen.

