**Quickstart – Segmentierung**

A. Segmentierung wird gebraucht:

* um die transkribierten Texte in Tokens zu unterteilen (Äußerungen, Worte, Pausen, nicht-phonologisches Material, Interpunktion) Bearbeiten > **Voreinstellungen… >** Segmentation
* für Visualisierungen Transkription > Transformation…
* zur Ausgabe von Äußerungslisten Transkription > Transformation… > HIAT Utterance list (HIAT)
* zum Exportieren von Wortlisten Transkription > Wortliste…
* zur Auszählung von Segmentketten Transkription > Transkription auszählen…
* für die Integration von Transkriptionen in ein EXMARaLDA-Korpus in Coma
* als Grundlage für die Suche im EXAKT

B. Zwei EXMARaLDA-Transkriptionsformate

|  |  |
| --- | --- |
| Basis-Transkription | Segmentierte Transkription |
| * Dateierweiterung: .exb[[1]](#footnote-1) * Die Elemente auf der Zeitachse und/oder Zeitpunkte dienen der Organisation von verschiedenen Ereignissen. * Es hat **keine** sprachliche oder anderweitig transkriptionsrelevante Bedeutung. * Selbst wenn für jedes Wort ein separates Ereignis erzeugt wird, werden diese Grenzen **nicht** als Wortgrenzen erkannt. | * Dateierweiterung: .exs * Informationen über die verschiedenen Einheiten in der Transkription werden **explizit** ausgedrückt, z.B. eine Äußerung, die aus mehreren Wörter besteht. * Erst nach der Segmentierung, können Wörter oder andere Einheiten, die durch verschiedene Transkriptionskonventionen markiert sind, erkannt werden. |

C. Segmentierung in EXMARaLDA

Der EXMARaLDA Partitur-Editor verfügt über eine **eingebaute** Segmentierung. Entweder eine für die am häufigsten gebrauchten *Transkriptionssysteme* (HIAT, DIDA, GAT, CHAT) und IPA, oder auch eine *generische* Segmentierungsfunktion.

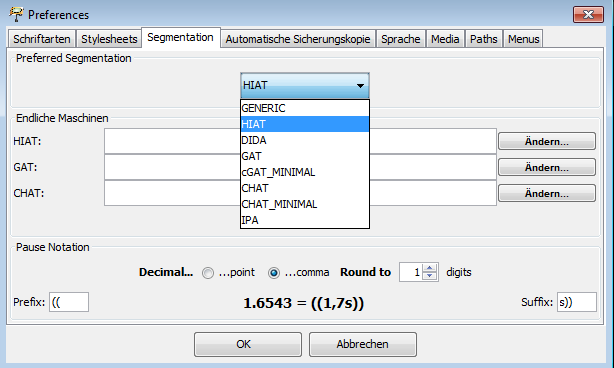
Allen Segmentierungstypen liegt das Prinzip zugrunde, dass die Segmentkette[[2]](#footnote-2) sich immer auf der obersten Ebene befindet, d.h. wenn eine Segmentkette versehentlich unterbrochen wird, wird auch die ganze Segmentierung fehlerhaft sein.



Aus dem Dropdown-Menü (1) Bearbeiten > Voreinstellungen… > Segmentation kann die bevorzugte Segmentierung ausgewählt werden. Danach mit OK bestätigen (2):

2

1

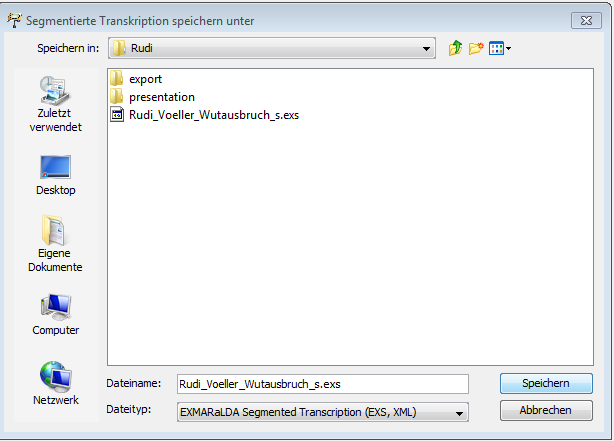


D. Exportieren segmentierter Transkriptionen

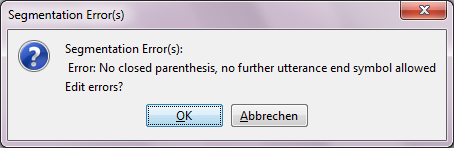
Segmentierte Transkriptionen können vom Partitur-Editor nicht gelesen werden. Die Basis-Transkriptionen sollten daher in diesem Prozess auf gar keinen Fall überschrieben werden. Um eine segmentierte Transkription zu erstellen (wie im Menü Bearbeiten > Voreinstellungen… > Segmentation festgelegt wurde), soll zunächst der Punkt Segmentierte Transkription exportieren… aus dem Menü Transkription ausgewählt werden. Sollte es *keine* Segmentierungsfehler geben, so kann die segmentierte Transkription gespeichert werden (1) – vorzugsweise mit einem Namen, der sich auf die Basis-Transkription bezieht und der mit dem Suffix \_s (2) erweitert wird:

2

1



Sollte die Segmentierung fehlerhaft sein, wird folgende Meldung angezeigt:

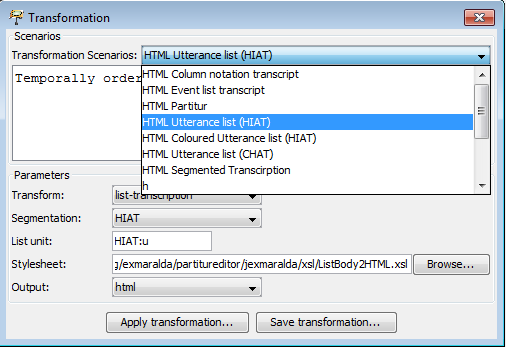


Durch einen Klick auf OK wird die Liste mit Segmentierungsfehlern eingeblendet.

Im Dokument “Quickstart correcting structure and segmentation errors” sind weitere Schritte zur Behebung solcher Fehler erläutert.

E. Visualisierungen

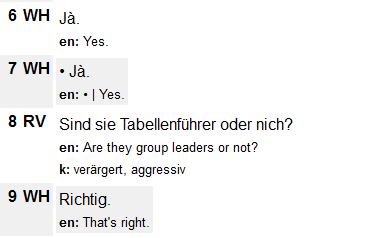
Aus einer segmentierten Transkription lassen sich verschiedene Visualisierungen erstellen. Für die Ausgabe (Output) kann ein völlig anderes Format als die im Partitur-Editor verwendete musikalische Partitur ausgewählt werden. Über Transkription > Transformation… kann die gewünschte Visualisierung, z.B. eine HIAT-Äußerungsliste (1) ausgewähltt werden:



1

Dann Apply Transformation… (2) wählen.

Die Resultate können unter Save transfromation… (3) gespeichert werden.



3

2

Die Visualisierung der Äußerungsliste weist nur wenig Ähnlichkeit mit der musikalischen Partitur auf[[3]](#footnote-3):

F. Wortlisten

Über Transkription > Wortliste… kann eine Wortliste generiert werden:



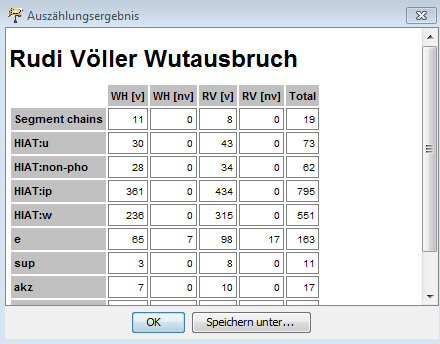
Durch Klicken auf das Feld Word (Wort) (1) oder Speaker (Sprecher) (2) kann diese Liste, gemäß des gewünschten Kriteriums, alphabetisch sortiert und das Resultat anschließend als eine .html-Datei gespeichert werden (Save as…).

2

Auch mithilfe einer Wortliste, können Rechtschreib- bzw. Tippfehler aufgedeckt werden[[4]](#footnote-4). Nachdem Sie die Fehler in der Transkription verbessert haben, drücken Sie auf Refresh.

1

H. Transkription auszählen



Eine Liste mit Häufigkeiten für verschiedene Arten von Segmenten (für den ausgewählten Segmentierungsalgorithmus) kann über Transkription > Transkription auszählen… erstellt werden. Die Auszählung kann über (Speichern unter…) gespeichert werden.

1. Die Erweiterungen .exb und .exs werden von den EXMARaLDA-Werkzeugen gebraucht, um das entsprechende Format zu erkennen; alle Dateien in EXMARaLDA sind XML-Dateien. [↑](#footnote-ref-1)
2. Segmentketten sind ununterbrochenen Serien von Ereignissen, die nur einem Sprecher gehören. [↑](#footnote-ref-2)
3. Die verschiedenen Stylesheets können auf [www.exmaralda.org](http://www.exmaralda.org) aus der Sektion „Downloads“ heruntergeladen werden. [↑](#footnote-ref-3)
4. Für weitere Informationen s. „Quickstart correcting structure and segmentation errors“. [↑](#footnote-ref-4)