

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

- Dateierweiterung: .exb¹
- 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.

Segmentierte Transkription

- Dateierweiterung: .exs
- Informationen über die verschiedenen Einheiten in der Transkription werden **explizit** ausgedrückt, z.B. eine Äußerung, die aus mehreren Wörtern 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.

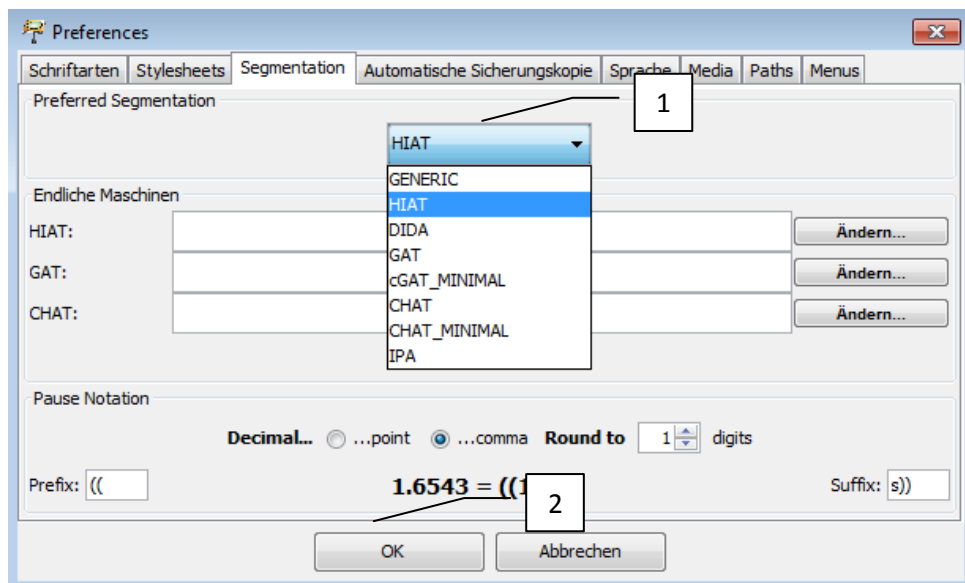
Allen Segmentierungstypen liegt das Prinzip zugrunde, dass die Segmentkette² sich immer auf der obersten Ebene befindet, d.h. wenn eine Segmentkette versehentlich unterbrochen wird, wird auch die ganze Segmentierung fehlerhaft sein.



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

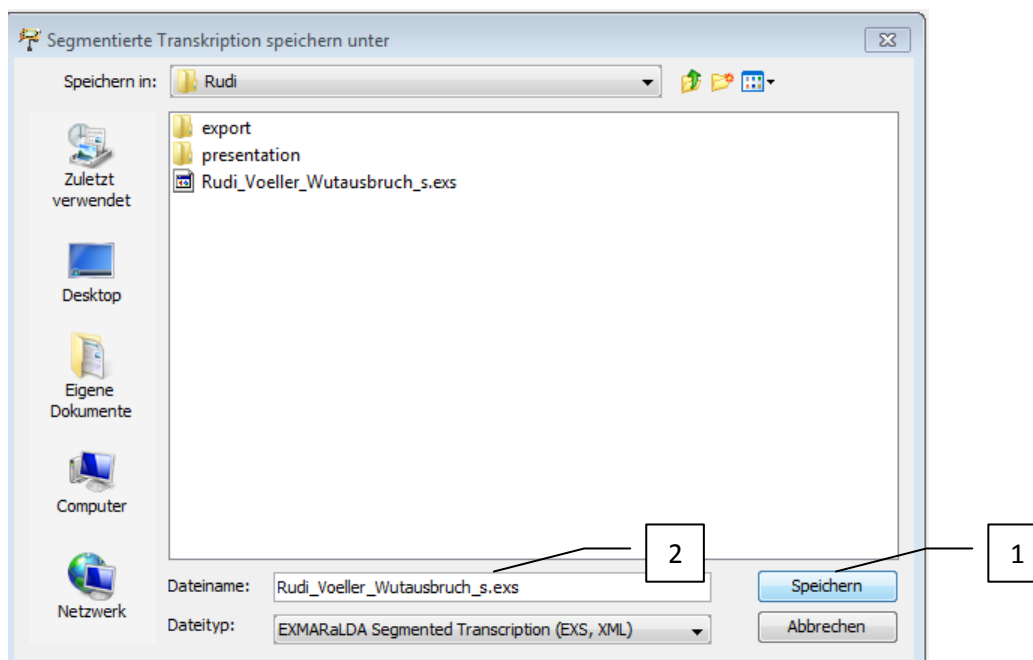
¹ Die Erweiterungen .exb und .exs werden von den EXMARaLDA-Werkzeugen gebraucht, um das entsprechende Format zu erkennen; alle Dateien in EXMARaLDA sind XML-Dateien.

² Segmentketten sind ununterbrochenen Serien von Ereignissen, die nur einem Sprecher gehören.



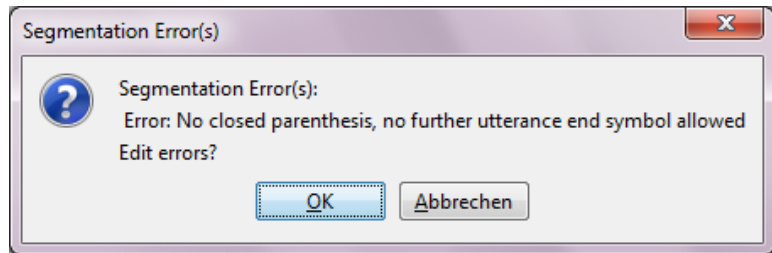
D. Exportieren segmentierter Transkriptionen

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:



Sollte die Segmentierung fehlerhaft sein, wird folgende Meldung angezeigt:

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



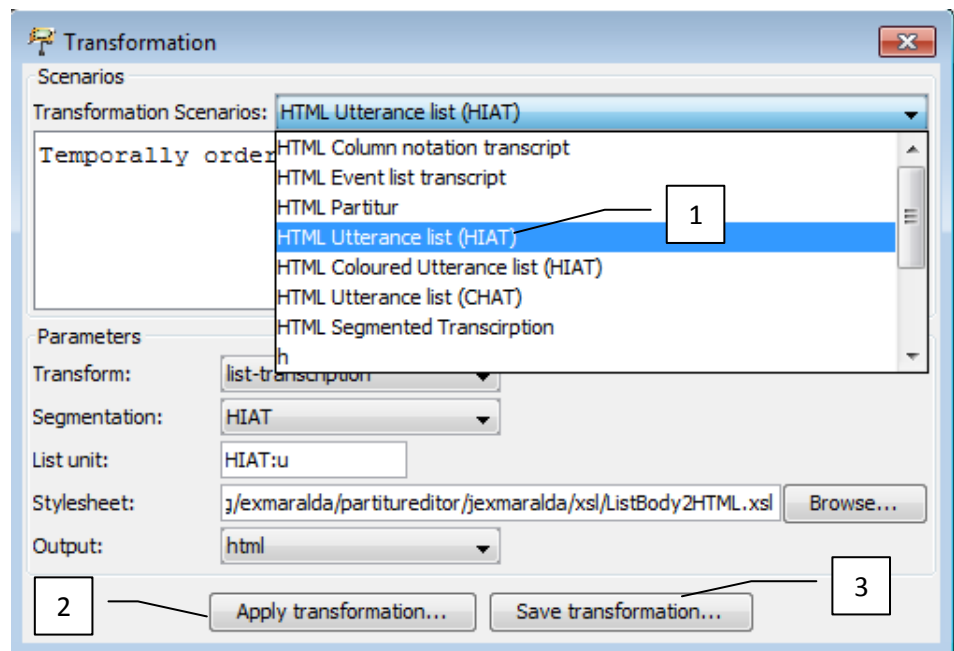
Im Dokument “Quickstart correcting structure and segmentation errors” sind weitere Schritte zur Behebung solcher Fehler erläutert.

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) ausgewählt werden:

Dann **Apply Transformation...** (2) wählen.

Die Resultate können unter **Save transformation...** (3) gespeichert werden.



6 WH Jà.

en: Yes.

7 WH • Jà.

en: • | Yes.

8 RV Sind sie Tabellenführer oder nicht?

en: Are they group leaders or not?

k: verärgert, aggressiv

9 WH Richtig.

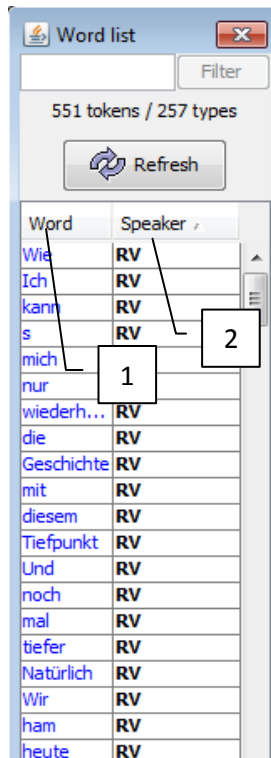
en: That's right.

Die Visualisierung der Äußerungsliste weist nur wenig Ähnlichkeit mit der musikalischen Partitur auf³:

³ Die verschiedenen Stylesheets können auf www.exmaralda.org aus der Sektion „Downloads“ heruntergeladen werden.

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...**).

Auch mithilfe einer Wortliste, können Rechtschreib- bzw. Tippfehler aufgedeckt werden⁴. Nachdem Sie die Fehler in der Transkription verbessert haben, drücken Sie auf **Refresh**.

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.

	WH [v]	WH [nv]	RV [v]	RV [nv]	Total
Segment chains	11	0	8	0	19
HIAT:u	30	0	43	0	73
HIAT:non-pho	28	0	34	0	62
HIAT:ip	361	0	434	0	795
HIAT:w	236	0	315	0	551
e	65	7	98	17	163
sup	3	0	8	0	11
akz	7	0	10	0	17

⁴ Für weitere Informationen s. „Quickstart correcting structure and segmentation errors“.