

Quickstart – Segmentation

A. Segmentation is needed:

- to divide the transcribed texts into tokens (utterances, words, pauses, non-phonological material, punctuation) **Edit > Preferences... > Segmentation**
- for visualizations **Transcription > Transformation...**
- to output utterance lists **Transcription > Transformation... > HIAT utterance list (HIAT)**
- to export word lists **Transcription > Word list...**
- to count segments **Transcription > Count segments...**
- to integrate the transcriptions into an EXMARaLDA corpus in Coma
- and as basis for EXAKT searches

B. Two EXMARaLDA transcription formats:

Basic transcription

- File extension: .exb¹
- The timeline items or time points are used to organize the different events
- There is **no** linguistic or otherwise transcription-relevant meaning
- Even if you create one event for each word, these boundaries **will not** be recognized as word boundaries.

Segmented Transcription

- File extension: .exs
- The information about different units in the transcription is expressed **explicitly**; it is partly organized according to these units, e.g. an utterance consisting of words.
- Only after segmentation **words** or other units denoted by the transcription conventions' different symbols **can be** recognized

C. Segmentation in EXMARaLDA

The EXMARaLDA Partitur-Editor comes with a **built-in** segmentation either for the *transcription systems* most commonly used (HIAT, DIDA, GAT, CHAT) and IPA, or there is also a *generic* segmentation.

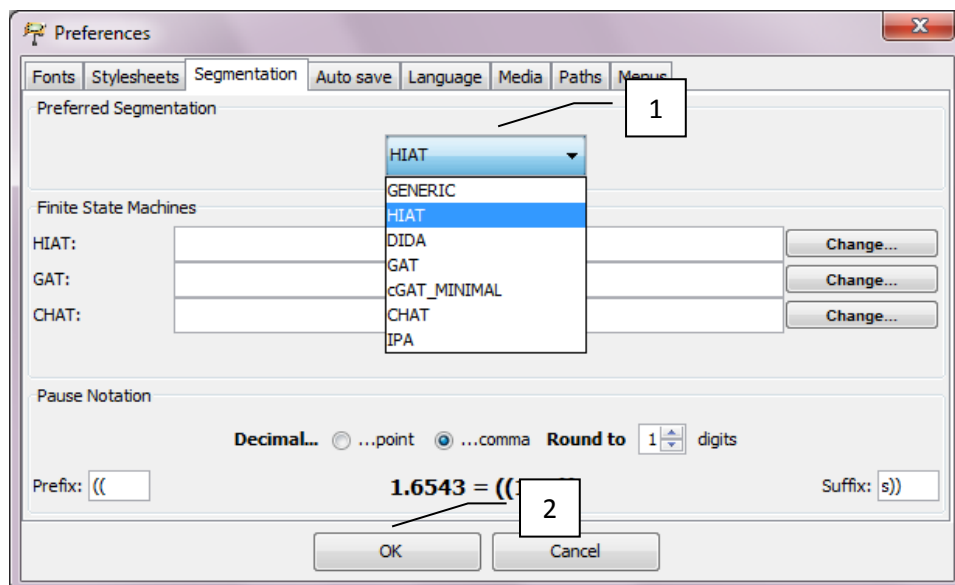
All segmentations use the notion of segment chains² as a top level, i.e. if a segment chain is erroneously interrupted, the segmentation will too be erroneous because the segment will be split.



Now, you can choose your preferred segmentation using the drop-down menu (1) at **Edit > Preferences > Segmentation**. Confirm your selection by clicking on **OK**:

¹ The extensions .exb and .exs are only there for the EXMARaLDA tools to recognize the appropriate file format; all EXMARaLDA files are XML files.

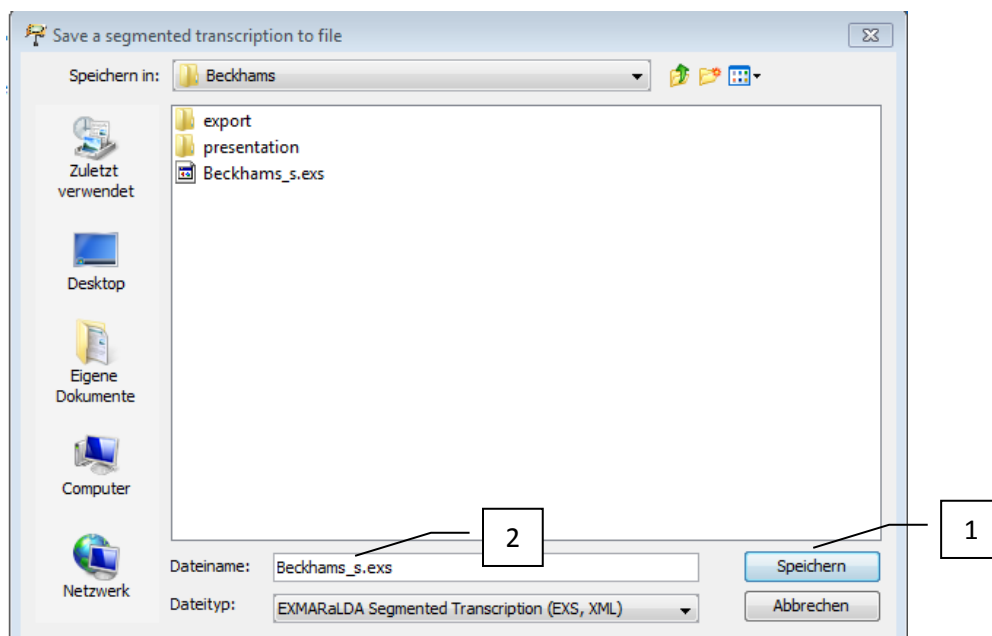
² Segment chains are the uninterrupted series of events belonging to one speaker.



D. Exporting segmented transcriptions

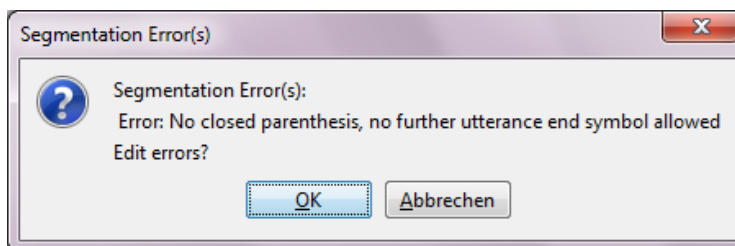
The Partitur-Editor cannot read the segmented transcription format. Therefore, you should pay attention not to overwrite the basic transcription file.

To create a segmented transcription with the segmentation algorithm specified at **Edit > Preferences > Segmentation**, choose **Export Segmented Transcription...** from the **Transcription** menu. If there are no segmentation errors, you can save the segmented transcription (1), preferably with a name related to the original basic transcription, e.g. by using the suffix `_s` (2):



If the transcription is not correct, you will receive an error message like this:

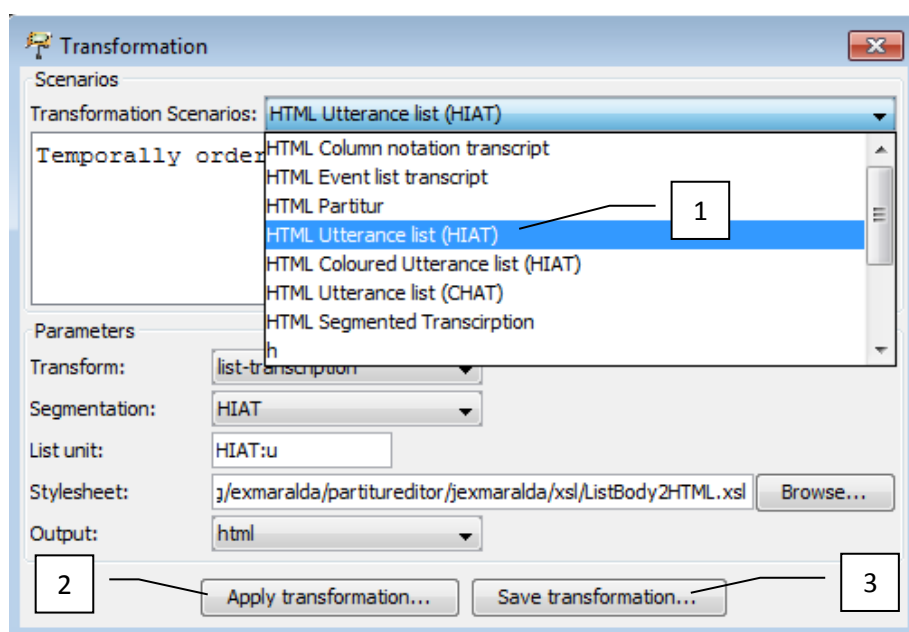
Click **OK** to arrive at the list of segmentation errors.



For more information on segmentation errors and how to deal with them, please consult the following document “Quickstart correcting structure and segmentation errors”.

E. Visualizations

It is possible to create different visualizations from one and the same segmented transcription. The output can be in a completely different format than the musical partiture used in the Partitur-Editor. Choose **Transcription > Transformation...** and click your preferred visualization for output, e.g. HIAT utterance list (1):



Then choose **Apply Transformation...** (2) and save your results via **Save transformation...** (3).

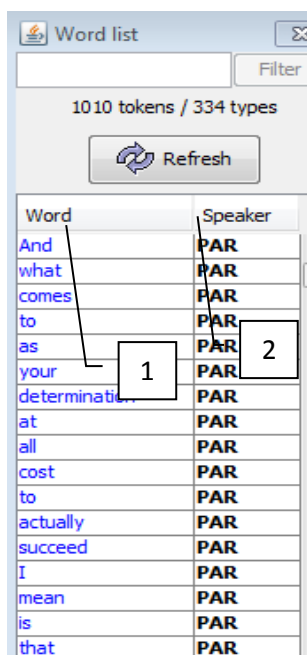
The visualized utterance list has a little resemblance to the musical score format:³

- 1 PAR** And what comes to as your determination at all cost to actually ((0,3s)) succeed.
- 2 VIC** Yeah
- 3 VIC** Mhm
- 4 PAR** I mean is that a/ sort of a message that you hope comes across to ((0,4s)) to kids?
- 5 PAR** Because a lot of kids think that people just become famous over night, don't they?

³ Please refer to the Download section at www.exmaralda.org for various style sheets for visualization.

F. Wordlists

To generate a word list, choose **Transcription > Word list...**:



By clicking on the header rows **Word** and **Speaker** you can sort the list alphabetically according to these criteria. Also with this list, you can check your transcription for the correct spelling and other typing errors.⁴ Having corrected the errors in the transcription, simply click on the **Refresh** button

You can save the word list as an .html-file at **Save as...**

H. Count segments

To generate a list of frequencies for the different kind of segments in the transcription with the chosen segmentation algorithm, choose **Transcription > Count segments...**

You can also save the results at **Save as...**

	PAR [v]	VIC [v]	DAV [v]	[nn]	Total
Segment chains	19	15	14	0	48
HIAT:u	37	42	34	0	113
HIAT:w	322	405	283	0	1010
HIAT:ip	485	661	433	0	1579
HIAT:non-pho	18	37	18	0	71

⁴ For more information on correcting these errors, see „Quickstart correcting structure and segmentation errors“.