## **Quickstart – Creating a corpus from a set of transcriptions**

## A. Preparation

The files of the EXMARaLDA Corpus on your hard drive should be organized as follows:

- 1. Media files should be located in the same folder or in a folder beneath the transcription they belong to.
- 2. Transcriptions that belong together (thematically or organizationally) should be held in the same directories.
- 3. The Coma file to be generated should be above these directories. In case of using the assistant "Create corpus from transcriptions...", this last point can be ignored as the file will automatically be located at the correct directory (s. B.).

The files in the Partitur-Editor should be standardized as follows:

- 1. Abbreviations in the speakertable should be identical for each speaker.
- 2. Attribute labels in the speakertable for the same attributes should be the same.
- 3. Attribute labels in the metadata for the same kind of attribute should be identical.

## B. Using the assistant

This assistant is used to create a Coma file from a set of EXMARaLDA-transcriptions that already have metadata stored in their transcriptionheads. It will lead you through the necessary steps:

 Open EXAKT and click on File > Create a corpus from transcriptions.... Enter the name for your corpus into the text field Name of corpus. Then click on the Browse...button and navigate to the directory above all directories that contain transcriptions to be added to the corpus. The filename will already be filled with the name you chose for thecorpus – if you want to change it, you can do it now. Continue by clicking next >.

Create corpus from	transcriptions					23
1. Coma file	Name of corpus, Coma file					
2. Transcriptions						
3. Segmentation	Name of corpus	unnamed				
4. Metadata assignment	Coma file			Browse		
5. Speaker						
6. Summary						
			cancel	< previous	next >	finish

2. In this step, you can select the transcriptions to be included into the corpus. You can (de-) select single files with the checkboxes in the first column. The checkboxes in the last column are not clickable - they signal whether the transcription is a basic (unchecked) or segmented (checked) transcription. You can choose to include only segmented transcriptions by clicking on the "deselect all" and then on the "select basic transcriptions"-button.

🧭 Create corpus from	transcriptions			X	When you are
1. Coma file	select transcriptions				
2. Transcriptions	Ind file	path		seg	done with
3. Segmentation 4. Metadata assignment 5. Speaker 6. Summary	ImplishTranslator_s.exs       ImplishTranslator_s.exs	C: \Users \Z2\Documents \Demo-Korr C: \Users \Z2\Documents \Demo-Korr	NUS NUS NUS NUS NUS NUS NUS		your selection, click <b>next &gt;</b> .
	Select segmented transcriptions 😨 select basic transcriptions	cancel	< previous   next >	finish	

3. In this step you can decide to create segmented from basic transcriptions while creating the corpus if you have not already done so. Further information about basic- and segmented transcription can be found in "Under-standing the basics of EXMARaLDA".

Coma file	Segmentation	
Transcriptions	segment transcriptions	
Segmentation	segmentation algorithm	default
	on segmentation errors	…cancel
Metadata assignment		@skip
Speaker		<ul> <li>use default segmentation</li> </ul>
Summary		✓write error list
	target	(ii) same directory
		new directory segmentedTranscriptions
	suffix	_s

When you are done with your segmentation settings, click next >.

With Metadata assignment you decide, how the metadata found in the transcriptions 4. (all fields found are shown in the second column of the table) will be used in the Comametadata-file.

metadata me.	Create corpus from transcriptions					
	1. Coma file	Meta	adata assignment			
	2. Transcriptions	use	field	target name	target	
If you are done	2 Commentation	V	@language-used-1	@language-used-1	communication	~
ii jou uit uone	5. Segmentation		comment	comment	communication	
making the metadata	4. Metadata assign		project-name	project-name	communication	
making the metadata			transcription-convention	transcription-convention	communication	
	5. Speaker	V	transcription-name	transcription-name	communication	
assignments, click		<b>V</b>	ud_Background information	Background information	communication	
0	6. Summary	<b>V</b>	ud_COMA-Communication-ID	COMA-Communication-ID	communication .	-
novt >		<b>V</b>	ud_COMA-Transcription-ID	COMA-Transcription-ID	communication	11
IICAL P.		<b>V</b>	ud_Code	Code	communication	
		<b>V</b>	ud_First name	First name	communication	
		V	ud_Gesprächstyp	Gesprächstyp	communication	
		<b>V</b>	ud_Last name	Last name	communication	
			ud_Profession	Profession	communication	4
		<b>V</b>	ud_Source	Source	communication	
		<b>V</b>	ud_Type of communication	Type of communication	communication	
		<b>V</b>	ud_c:Aufnahmedatum	c:Aufnahmedatum	communication	
		V	ud_c:Gesprächstyp	c:Gesprächstyp	communication	
		<b>V</b>	ud_c:Quelle	c:Quelle	communication	_
			ud. c:Transkribent	c•Transkribent	communication	-
		commu	unication name from transcription-name			•
					cancel corevious next > finish	$\neg$

5. On the "**Speaker**-screen" you get a list of all speakers found in the speakertables of the found transcriptions. Now you can (de-)select which of the speakers should be included and via which metadata-field (use drop-down-menu) in the Coma-metadata.

1. Coma file       Speaker         2. Transcriptions       use       @abbreviation         3. Segmentation       Ø       AnR         4. Metadata assignment       Ø       AW         Ø       AW       Ø         Ø       AW       Ø         Ø       AW       Ø         Ø       AW       Ø         Ø       BS       Ø         Ø       C       Ø         Ø       C       Ø         Ø       GU       Ø         Ø       GU       Ø         Ø       HG       Ø         Ø       HM       Ø         Ø       MOD       Ø         Ø       P       Ø         Ø       P       Ø         Ø       P       Ø         Ø       P       Ø         Ø       P       Ø         Ø       P       Ø         Ø       P       Ø         Ø       P       Ø         Ø       P       Ø         Ø       P       Ø         Ø       P       Ø         Ø       Ø       Ø	×
2. Transcriptions       use       @abbreviation         3. Segmentation       ✓       ANR         4. Metadata assignment       ✓       AW         5. Speaker       ✓       C         6. Summary       ✓       GL         ✓       GL       ✓         ✓       GW       ✓         ✓       MOD       ✓         ✓       NN       ✓         ✓       P       ✓         ✓       P       ✓         ✓       P       ✓         ✓       P       ✓         ✓       P       ✓         ✓       P       ✓         ✓       P       ✓         ✓       P       ✓         ✓       P       ✓         ✓       P       ✓         ✓       P       ✓         ✓       P       ✓         ✓       P       ✓         ✓       P       ✓         ✓       P	
J. Segmentation       ▼       ANR         4. Metadata assignment       ▼       ANR         5. Speaker       ▼       C         Ø       C       ●         Ø       AV       ●         S. Speaker       ▼       C         Ø       C       ●         Ø       Q       C         Ø       C       ●         Ø       G       ●         Ø       G       ●         Ø       HM       ●         Ø       HM       ●         Ø       HM       ●         Ø       NN       ●         Ø       PAR       ●         Ø       Pace =       ●         Ø       Pab	
> Segment tool         ▼         ANR           4. Metadata asignment         ✓         AW           S. Speaker         ✓         C           Ø         DAV         ✓           S. Summary         ✓         GL           Ø         GW         ✓           Ø         MG         ✓           Ø         MG         ✓           Ø         MG         ✓           Ø         MOD         ✓           Ø         NN         ✓           Ø         PAR         ✓           Ø         PAR         ✓           Ø         Park         ✓           Ø         Sone         ✓           unque speaker distriction:         @abbreviation         © eaver           Ø10           Ø         Ø	-
4. Metadata assignment         ✓         AW           5. Speaker         ✓         BS           ✓         C            S. Speaker         ✓         C           ✓         C            ✓         C            ✓         C            ✓         C            ✓         C            ✓         C            ✓         C            ✓         C            ✓         MagR            ✓         NN            ✓         P            ✓         PAR            ✓         PAR            ✓         PaR            ✓         Particitant         @abbreviation           Inique speaker distriction:         @abbreviation         @abbreviation           @atbreviation         @atbreviation         @atbreviation	
S. Speaker         V         BS           Ø         C         O           Ø         DAV         O           Ø         DAV         O           Ø         G         O           Ø         G         O           Ø         G         O           Ø         HM         O           Ø         MOD         O           Ø         MOD         O           Ø         NN         O           Ø         PAR         O           Ø         PAR         O           Ø         REP         O           Ø         REP         O           Ø         Sone         O           Ø         O         O           Ø         Babrevision         Poeudo/name           Ø         O         O	
b. Spearer         Ø         C           6. Summary         Ø         DAV           Ø         GL         Ø           Ø         GW         Ø           Ø         HM         Ø           Ø         HM         Ø           Ø         MOD         Ø           Ø         MOD         Ø           Ø         MOD         Ø           Ø         P         Ø           Ø         P AR         Ø           Ø         PawW         Ø           Ø         PawW         Ø           Ø         PawW         Ø           Ø         RP         Ø           Ø         RP         Ø           Ø         Rev         Ø           Ø         Sealbreviation         Ø           Ø         Ø         Ø           Ø         Ø         Ø           Ø         Ø         Ø           Ø         Ø         Ø           Ø         Ø         Ø           Ø         Ø         Ø           Ø         Ø         Ø           Ø         Ø         Ø	
6. Summary	
✓     GW       ✓     HG       ✓     HG       ✓     MoD       ✓     MoR       ✓     MagR       ✓     MagR       ✓     Pa       ✓     Pa       ✓     PaWW       ✓     PaWW       ✓     REP       ✓     RV       ✓     Sone       unique speaker distinction:     Gabbreviation       Ø1.0     Image: Construction       Ø1.0     Image: Construction	
Image: Provide the second	=
✓     HM       ✓     MOD       ✓     Mod       ✓     MagR       ✓     NN       ✓     P       ✓     Park       ✓     Park       ✓     Park       ✓     REP       ✓     RV       Ø     Rep       Ø     Rep       Ø     Repression       Ø     Babbreviation       Ø     Gabbreviation       Ø     Ø       Ø     Ø       Ø     Ø	
✓     MoD       ✓     MagR       ✓     NN       ✓     P       ✓     Park       ✓     PawW       ✓     PawW       ✓     REP       ✓     RV       ✓     SonR       unique speaker distinction:     @abbreviation       Ø10     ■       Ø10     ■       Ø10     ■	
✓     MagR.       ✓     NN       ✓     P       ✓     PaR       ✓     PaWW       ✓     REP       ✓     Rev       ✓     Sene       unique speaker distinction:     Gabbreviation       Ø10     Ø10       Ø10     Ø10	
✓     NN       ✓     P       ✓     Par       ✓     Par       ✓     Par       ✓     RP       ✓     RP       ✓     RV       ✓     Rone       unique speaker distinction:     @abbreviation       Ø10     Ø10       Ø10     Ø10	_
✓     PAR       ✓     PawW       ✓     PawW       ✓     REP       ✓     RV       ✓     SonR       unique speaker distinction:     @abbreviation       @abbreviation     @abbreviation       @abbreviation     @abbreviation       @abbreviation     @abbreviation       @abbreviation     @abbreviation       @abbreviation     @abbreviation       @abbreviation     @abbreviation	
✓     Pack       ✓     Pack       ✓     Pack       ✓     REP       ✓     RV       ✓     Son®       unique speaker distinction:     Gabbreviation       Ø10     Ø10       Ø10     Ø10	
V     Participant       V     REP       V     RV       V     SonR       Unique speaker distinction:     @abbreviation       @abbreviation        @abbreviation        @abbreviation        @abbreviation        @abbreviation        @abbreviation        @abbreviation        @abbreviation        @abbreviation	
✓     RV       IZ     Son8.       unique speaker distriction:     @abbreviation       Øabbreviation     ✓	
Unique speaker distinction: @abbreviation	
unique speaker distinction: @abbreviation	
Babbreviation Babbreviation Gillo Gillo Gillo	-
©10 ©10	^
comment	E
Connert Connert	
id comment	
UG_AGE ID	
ud Aushidung shulish ud Aushidung shulish	
ud Beruf	

When you are done with your speaker selection, press **next >**.

6. The next screen summarizes your choices on one screen. If you realize errors in the setup, just click on the "< previous"-button to correct the corresponding setting. When you are comfortable with your settings, click the **finish**-button.

por create corpus ironi	transcriptions			
1. Coma file	Summary			
2. Transcriptions	project-name->project-name			
3 Segmentation	ud_Profession->Profession			
5. Segmentation	transcription-name->transcription-name			
4. Metadata assignment	ud_c:Vorgeschichte->c:Vorgeschichte			
	ud_Type of communication->Type of communication			
o. Speaker	ud_c:project-name->c:project-name			
6. Summary	ud_c:Quelle->c:Quelle			
	comment->comment			
	ud_c:Aufnahmedatum->c:Aufnahmedatum			
	ud_Code->Code			
	Transcriptions:			
	Recordings:			
	Unique speakers selected:24			
	Segmentation options:			
		cancel	< previous	next > finish

If all goes well, the Coma-file is written and opened as the active document inside Coma right away.

r .		< Search			Filter		
and the second second	Tommercations	1		Gearth	instant of filters . has	inenker .	
				[			
munication					S Sple	Var	
Name	Yar.				HG		
Anne Will: Halbes Wahir	.7				SonB		
Rud Voller Wutausbruch	7	1			85		
Serif Issi: Expertenausk					AW		
Interview with Nguyen	2				- 11%		
Gasperini: Caldo	2				Land Land		
English Translator	-				MOD		
Szymunindjewski Raddaran	-				VIC		
eovare .					AR		
					GW		
					MN		
					REP		
					DAV		
					10		
					PAR		
					nom		
			Nechricht	ten Xee	P		
					PawW		
			Come lie wellen		a		
					π		
					MagR		
					SzyM		