Processing Morphological Transcriptions
ELAN and FLEx

Authors:

Paulien Veenstra (University of Göttingen)

Jana Bajorat (Humboldt-Universität zu Berlin)

Marie Benzerrak (University of Göttingen)

Discussion:

Diana Kakashvili (Ivane Javakhishvili Tbilisi State University)

Stavros Skopeteas (University of Göttingen)

version 1.0.0

Contents

1. General

Step 1	Step 2	Step 3
ELAN ⟶	FLEx ⟶	ELAN
Transcribe the spoken data in ELAN sentence-wise (including subordinate or coordinate clauses). The transcription is only basic phonological/orthographic transcription (without morpheme boundaries and glosses). You, then, will have two main layers: One layer with the sentence in the vernacular language and a layer of a free translation of that sentence.	FLEx automatically divides the sentence into words and adds two layers, a layer of morphemic transcription (i.e., morpheme boundaries) and a layer of glosses.	Import the FLEx file into ELAN. The data is now aligned with sound again. It can be integrated in database systems and it can be used to add further annotation layers for looking at specific phenomena, e.g., information structure.

Additional information:

- Create a template or set up an ELAN file before you annotate (with consultants).
    - A template contains the basic tier structure. Consequently, you only have to set up the ELAN file once, and this file can be used for all recordings.
- In the field you may have to work with ELAN and FLEx simultaneously because the dependencies in ELAN do not allow glossing on the word-level. Therefore, it may be practical to directly gloss/create the lexicon in FLEx.
- You can use note tiers if you want to, e.g., gloss in ELAN. If there is more than one note per unit in FLEx and these are imported into ELAN then the notes get merged into one single tier automatically.
- It is not possible to choose which layers/tiers you want to export out of FLEx into ELAN.
- If the guidelines do not work check your language settings in ELAN as well as FLEx and the tier names, this might solve the problem.

2. Setting up an ELAN File

Target Structure in ELAN:

| Tier Number	Tier Name	Content	Tier Type	Stereotype	Dependency
1	interlinear-title-en(g)	title	default	none	none
2	phrase-segnum-en(g)	sentence numbers	phrase	none	none
3	phrase-txt-arh	transcription	phrase-item	symbolic association	phrase-segnum-en(g)
4	free_translation-gls-en(g)	free translation	text-is-translation	symbolic association	phrase-segnum-en(g)
(5)	note-note-en(g)	comments and notes	note	symbolic association	phrase-segnum-en(g)
(6)	gloss_note-note-en(g)	glosses and/or transcription	note	symbolic association	phrase-segnum-en(g)
(7)	*language*-note-en(g)	translation in another language than the language of analysis	note	symbolic association	phrase-segnum-en(g)
(8)	*alphabet*-note-en(g)	usage of a different alphabet, orthography or, e.g., IPA	note	symbolic association	phrase-segnum-en(g)

2.1 Creating a File

2.1.1 Open ELAN. Click on *File > New*.
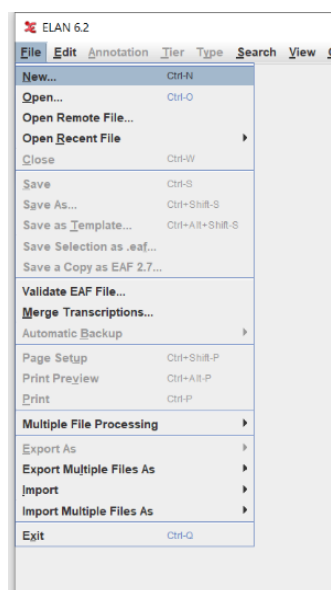
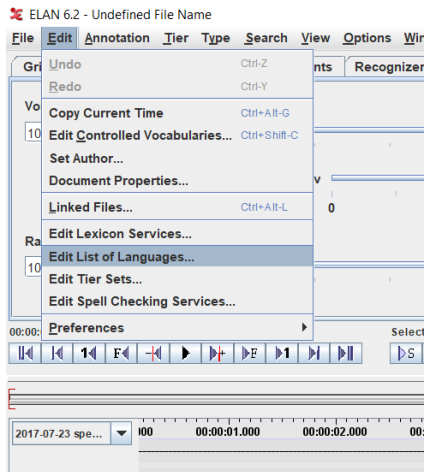## 2.1.2 Select an audio file. Click on >> and then on *OK*.
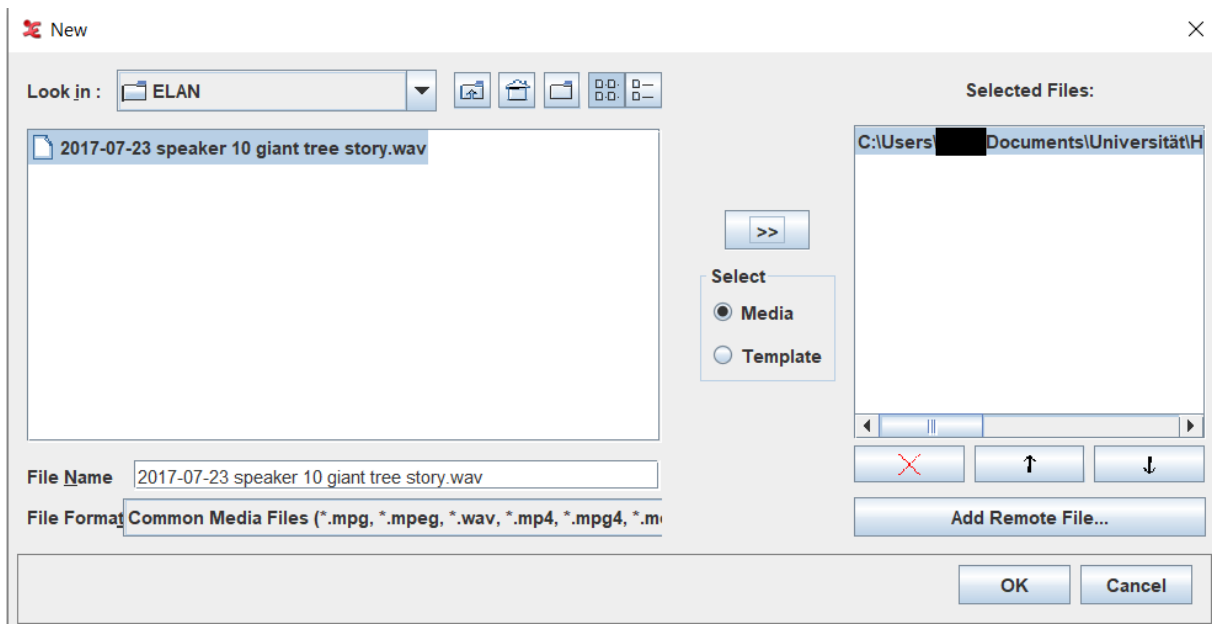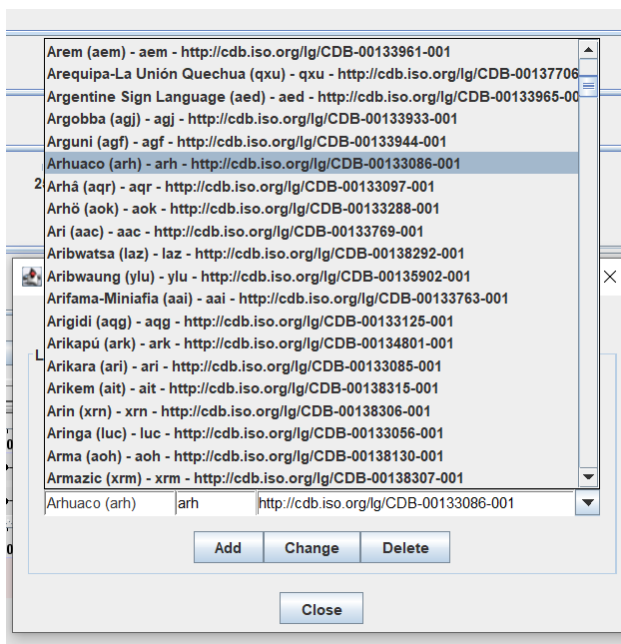

The result:
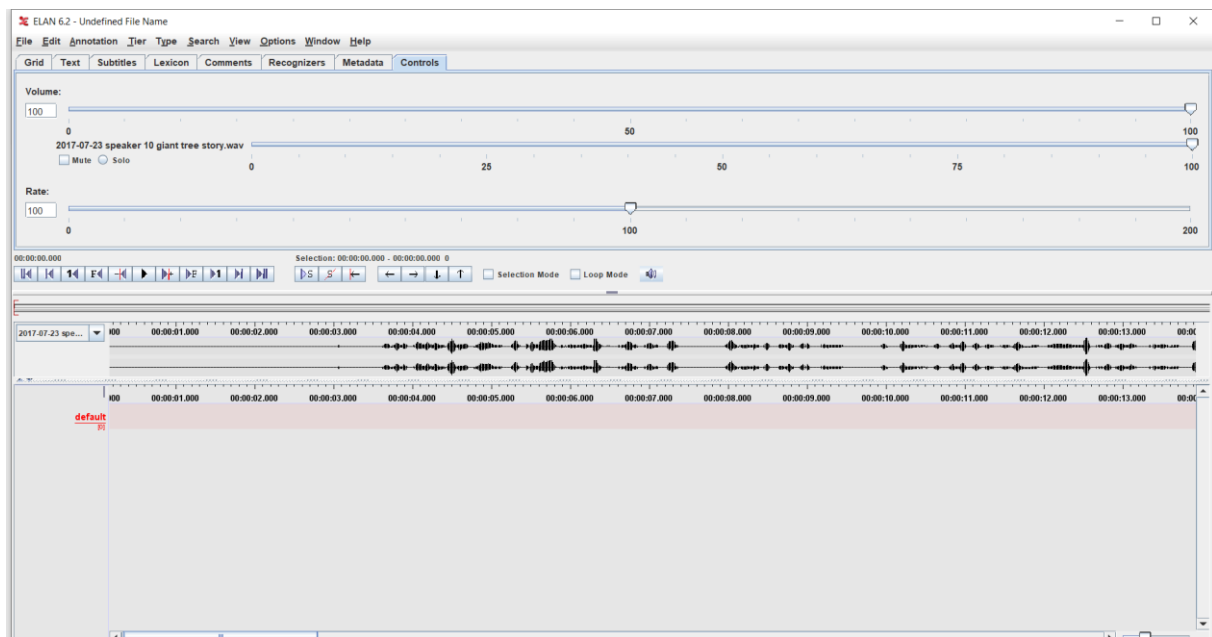
2.2 Language Settings

2.2.1 Before you can start annotating, you have to select the language you are working with (and the meta language, if necessary). Go to **Edit > Edit List of Languages** and add the languages you are working with.
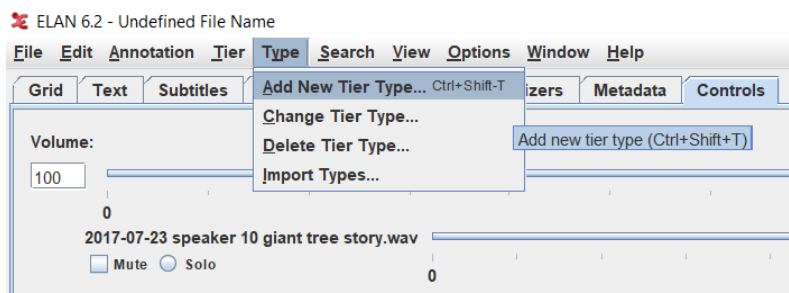

2.2.2 Click on the arrow, select a language, and click on *Add*. Close the window.



2.3 Setting up Tier Types

2.3.1 Click on the tab **Type > Add New Tier Type**.

2.3.2 There are two options: 2.3.2.1 create your own Tier Types or 2.3.2.2 import Tier Types from an already existing file.

2.3.2.1 You can **add your own tier types**. It is recommended to use the following names and stereotypes:

Type Name	Stereotype
phrase	none
phrase-item	symbolic association
text-is-translation	symbolic association
note	symbolic association

Add a **Type Name** and a **Stereotype**.
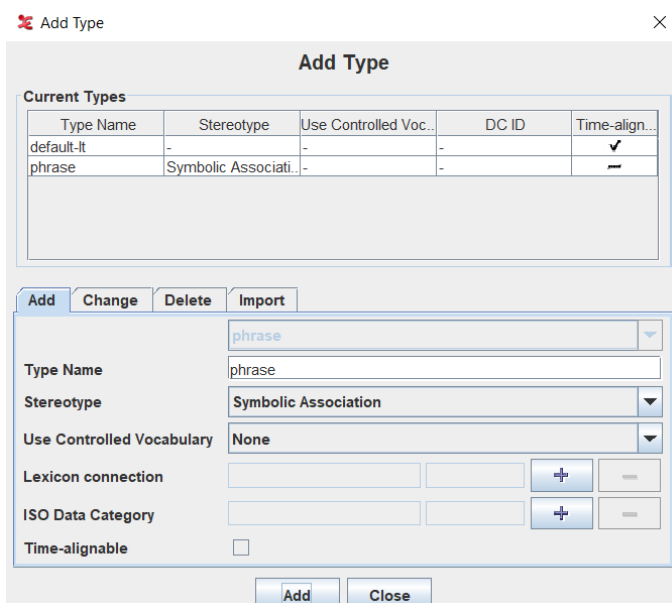
You can also add your own tier types. If you want to do so, you must keep the following in mind:

- It is best when you stick to the notions phrase, word, morph, phrase-item, word-item, morph-item or other notions that are embedded in ELAN and FLEx.
- What stereotype you choose depends on the annotations and the tier dependencies. These stereotypes define the kind of dependency a tier has regarding its parent. If a tier does not have a parent, no stereotype needs to be selected.
- If you are working with more tiers (than suggested here) and want to create more than one parent tier, you need to work with a more complex structure and, therefore, will need to adjust the tier types and/or add more tier types. Some names could be *word* or *morph*.

The **stereotypes** are defined as followed:

- **None**: "The annotation on the tier is linked directly to the time axis, i.e., the annotation is entered on an independent tier. Two annotations cannot overlap."
- **Time Subdivision:** "The annotation on the parent tier can be sub-divided into smaller units, which, in turn, can be linked to time intervals. Note that there are no time gaps allowed, i.e., the smaller units have to immediately follow each other. E.g., an utterance

transcribed on a parent tier can be sub-divided into words – each of which is then linked to its corresponding time interval. Note: Annotations on such tiers are time-alignable. They differ from annotations on independent tiers in that they are assigned to an interval that is contained within the interval of their parent annotation."

- ***Symbolic Subdivision:*** "Similar to Time Subdivision, except that the smaller units cannot be linked to a time interval. E.g., a word on a parent tier can be sub-divided into individual morphemes (which are not linked to a time interval)."
- ***Included In***: "All annotations fall within the borders of the parent tier. However, there can be gaps between the child annotations. E.g., a sentence with a silence can be split up into words while the silence corresponds to a gap in the child annotations (i.e. the separate words)."
- ***Symbolic Association:*** "The annotation on the parent tier cannot be sub-divided further, i.e., there is a one-to-one correspondence between the parent annotation and its referring annotation. E.g., one sentence on a parent tier has exactly one free translation. Or one word has exactly one gloss."

https://www.mpi.nl/corpus/html/elan/ch05s01.html (Accessed: 2021-11-19)

2.3.2.2 To ***import*** an existing ELAN file go to the tab ***Import*** and select a file. Click on *Import*.
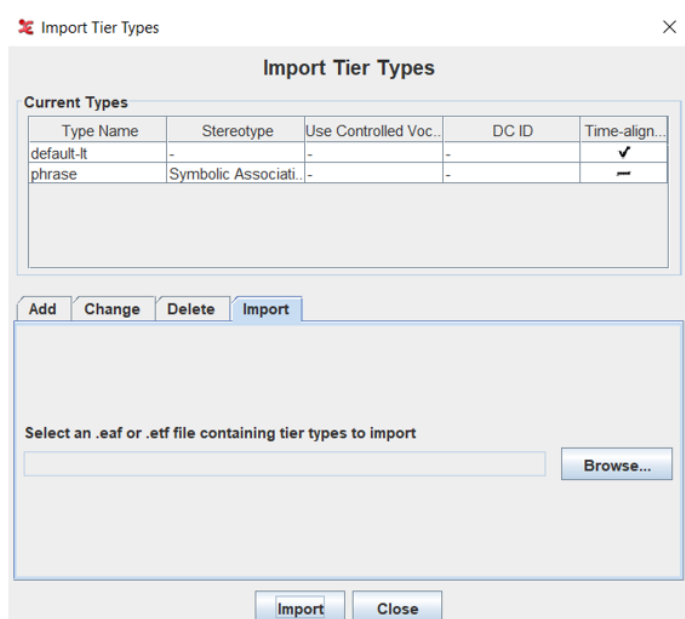
2.4 Tier Names

- Tier names consist of a ***name***, a ***FLEx type***, and an ***ISO code*** that should be separated by a hyphen, e.g., phrase-txt-arh.
- ***Name*** your tiers according to their kind of content, e.g., phrase, word, morph.
    - *Phrase* contains transcriptions (sentence-by-sentence).
    - *Word* contains the same material tokenized by words.
    - *Morph* includes morphemic boundaries and glosses.
- Examples of ***FLEx types***:
    - txt – text
    - text-is-translation – translation
    - gls – gloss
    - ps – part of speech

- o note – note
- The ***ISO code*** usually consist of three letters but depending on your ELAN-version and the ISO codes that you selected in step 2.2, a few languages can receive the ISO code 639-1, i.e., a code consisting of two letters, e.g., English - *en.* These languages are automatically put into the language list by ELAN, i.e., they do not need to be selected like the languages with the ISO code 639-3. Make sure the ISO codes you use match the ISO codes chosen in step 2.2.

The following names are recommended:

Tier Number	Tier Name	Content
1	interlinear-title-en(g)	title
2	phrase-segnum-en(g)	sentence numbers
3	phrase-txt-arh	transcription
4	free_translation-gls-en(g)	free translation
(5)	note-note-en(g)	comments and notes
(6)	gloss_note-note-en(g)	glosses and/or transcription
(7)	*language*-note-en(g)	translation in another language than the language of analysis
(8)	*alphabet*-note-en(g)	usage of a different alphabet, orthography or, e.g., IPA

The note tiers are optional and the amount of note tiers depends on your needs. In the table above you find some suggestions for the usage of note tiers. The italics in the tier names 7 and 8 can be changed to the name of the language or alphabet.

2.5 Setting up Tiers

2.5.1 Go to ***Tier > Add New Tier***.
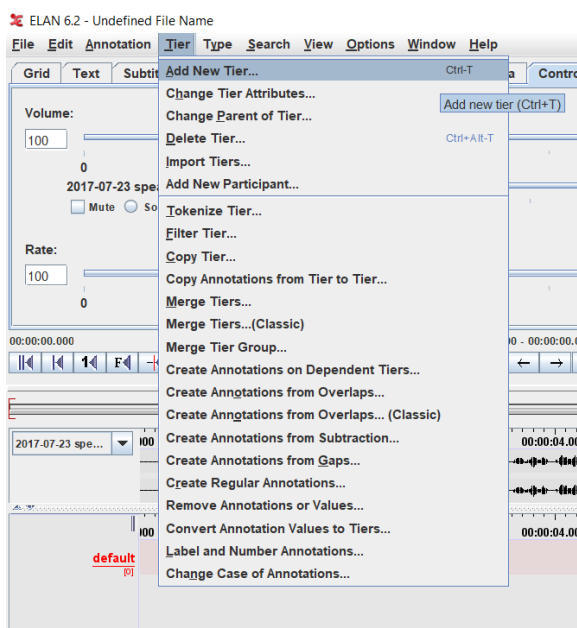
2.5.2 The **first tier** is an **interlinear-title-en(g)** tier (the language depends on the language you work with). Add the **tier name**, the **content language**, and further information that you find helpful. Nothing needs to be changed regarding parent tier and tier type. Click on *Add*.

Note that if you work with the **ISO code eng**, the selected content language looks as follows:

2.5.3 The **second tier** is called **phrase-segnum-en(g)** (the language depends on the language you work with). Add the **name**, the **tier type**, and the **content language**. Participant and annotator are optional. Click on *Add*.

2.5.4 The ***third tier*** is the ***phrase-txt-arh*** tier (transcription in the vernacular language). Here you have to add the **tier name**, **parent tier**, **tier type** and **content language**. Participant and annotator are optional. Click on *Add*.

2.5.5 The ***fourth tier*** is the ***free_translation-gls-en(g)*** tier. Note that it is crucial that you use the ***FLEx type gls*** in the tier name, otherwise the free translation will not appear in the right slot in FLEx. Add the **tier name**, **parent tier**, **tier type** and **content language**. Participant and annotator are optional. Click on *Add*.

2.5.6 The following tiers are *optional note tiers*. You can add as many as needed. The first part of the tier name must be unique, e.g., the table above and below with suggestions. Select the **parent tier phrase-segnum-en(g)**, the **tier type note**, and the **content language en(g)**. Participant and Annotator are optional. Note: you should set the ISO code to the analysis language (that you have selected in FLEx), only then the notes appear in the note slot in FLEx.[1]

---

[1] It is also possible to set the ISO code to the vernacular language, but then you have to change some settings in FLEX because this note slot is usually hidden.

2.5.7 Overview over the tier structure:

Tier Number	Tier Name	Content	Tier Type	Stereotype	Dependency
1	interlinear-title-en(g)	title	default	none	none
2	phrase-segnum-en(g)	sentence numbers	phrase	none	none
3	phrase-txt-arh	transcription	phrase-item	symbolic association	phrase-segnum-en(g)
4	free_translation-gls-en(g)	free translation	text-is-translation	symbolic association	phrase-segnum-en(g)
(5)	note-note-en(g)	comments and notes	note	symbolic association	phrase-segnum-en(g)
(6)	gloss_note-note-en(g)	glosses and/or transcription	note	symbolic association	phrase-segnum-en(g)
(7)	*language*-note-en(g)	translation in another language than the language of analysis	note	symbolic association	phrase-segnum-en(g)
(8)	*alphabet*-note-en(g)	usage of a different alphabet, orthography or, e.g., IPA	note	symbolic association	phrase-segnum-en(g)

2.5.8 The tiers have been created. Go to the tier names and click on the right mouse button and click on **Sort Tiers > Sort by Hierarchy**. Now the tiers appear in a **dependency hierarchy**.
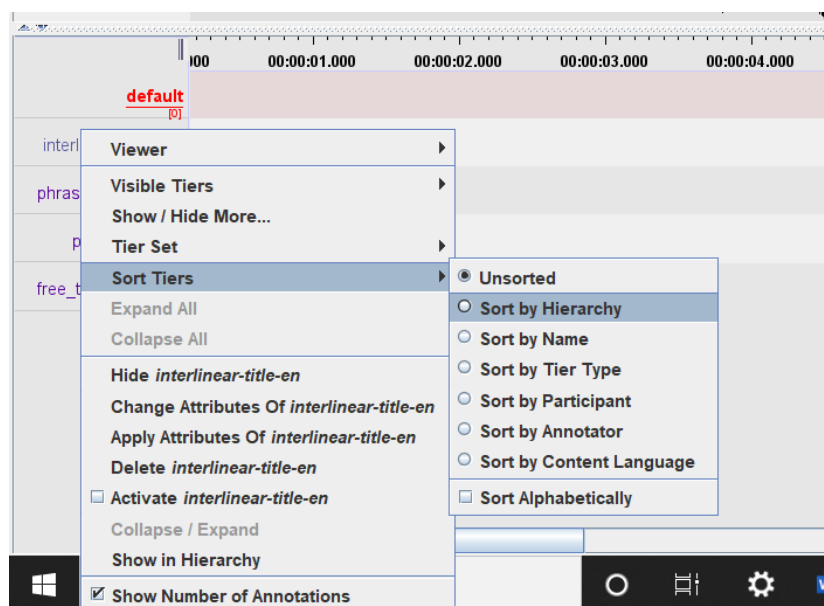
The result (without note tiers):
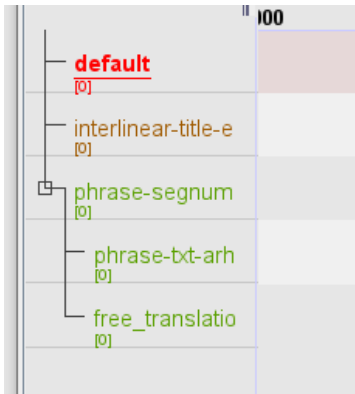
2.5.9 ***Delete the default tier***. Go to the name default and click on the right mouse button. Select ***delete default*** and follow the next instructions.[2]
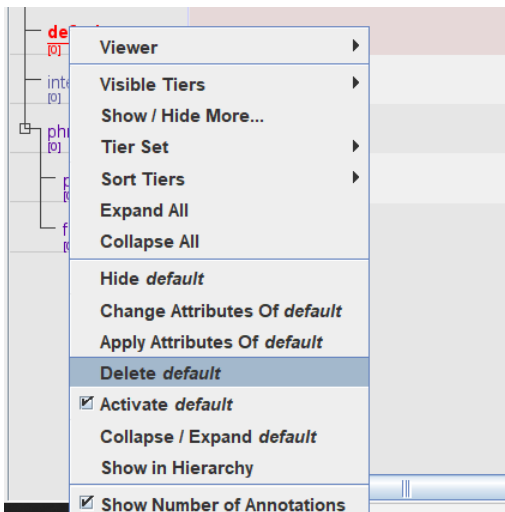
The result (with two note tiers):

_____
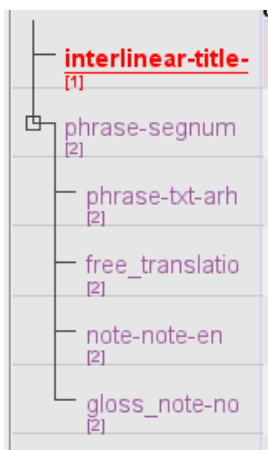
[2] Another option is changing the tier attributes of the default tier. You could name this tier interlinear-title-en(g) and add the tier attributes by clicking on **Tier > Change Tier Attributes**.

2.6 How to Create a Template

2.6.1 Creating a template file involves two steps:

1. Open a new ELAN file and create tier types and tiers as described above.
2. Then go to *File > Save as Template*.

2.6.2 Creating a new ***ELAN-project based on a template file***.

When creating a new ELAN-project you want to base on a template file, you must ***select*** two files: ***the media (or audio) file*** you want to transcribe and ***the template file*** you want to use. How to select the media file is also described above (step 2.1.2).
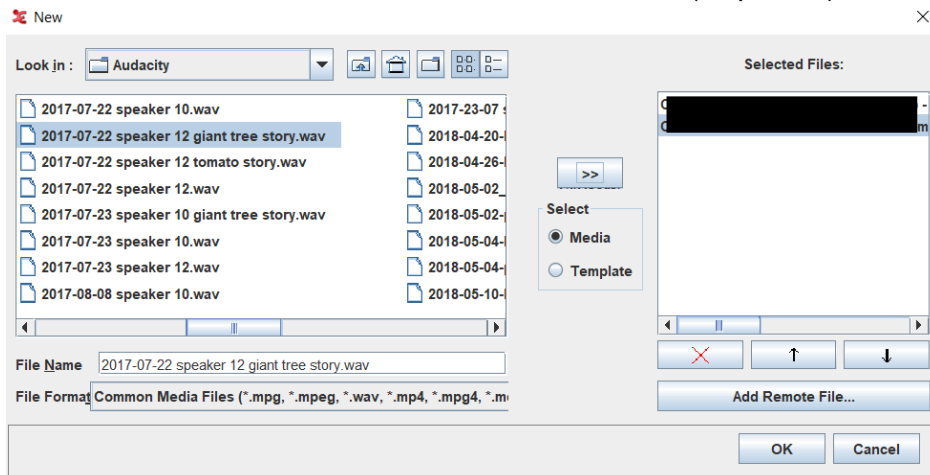
Then, additionally, you must select the template file in the same manner: Click on template, select the template file, and then, click on the arrows above *Select* in order to add the file to the selected files in the white box on the right side.
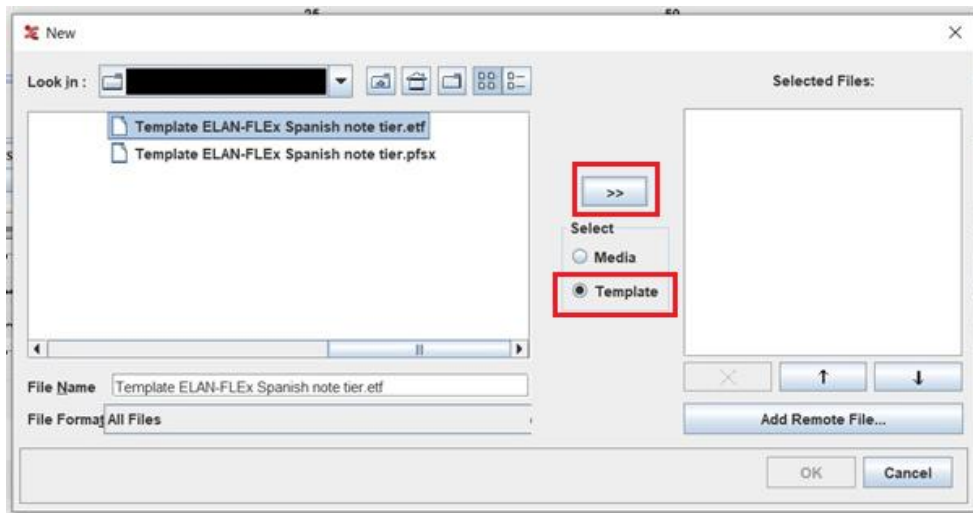
When you have selected both the template file and the media file (the order does not matter), click on *OK*.

2.7 Annotations

2.7.1 The *first annotation* is an annotation on the *interlinear-title-en(g)* tier. The annotation ranges from the beginning of the recording to the end. *Select the whole recording with your mouse*, the marked sequence will be *blue*. Then click on *Annotation > New Annotation Here or click twice* and add the title of the recording.
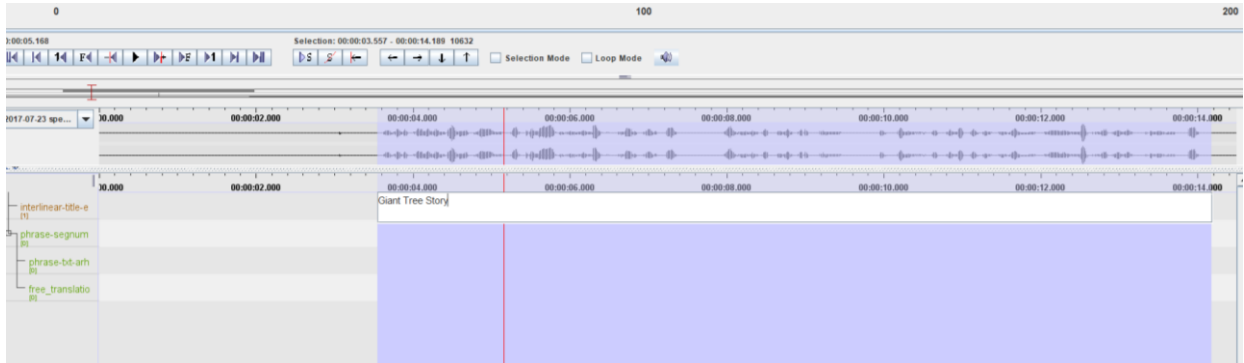
2.7.2 The *second tier (phrase-segnum-en(g))* consists of *annotations that are the length of a sentence*. Repeat the same process and add *numbers* from 1 upwards, i.e., the first sentence is one annotation and contains the number 1, the second sentence is the next annotation and contains the number 2 etc.
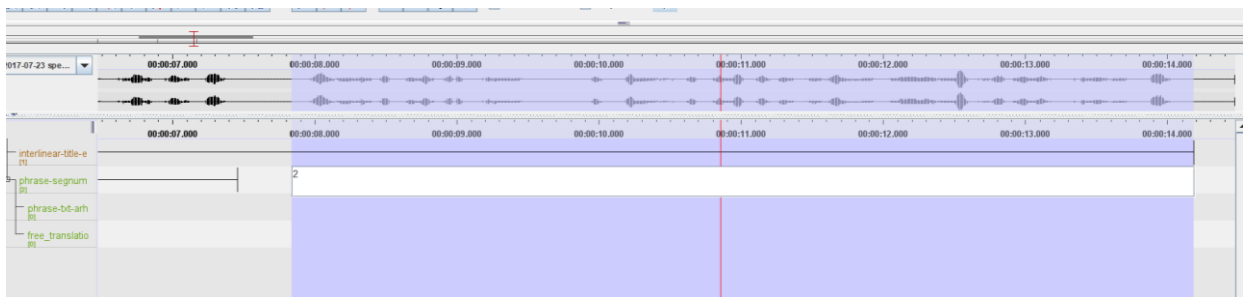
2.7.3 Creating Annotations on Dependent Tiers

2.7.3.1 Go to *Tier > Create Annotations on Dependent Tiers*.
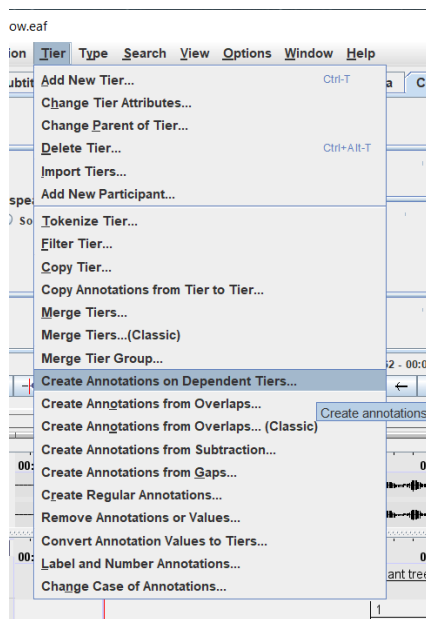
2.7.3.2 Select the **parent tier**, i.e., phrase-segnum-en(g). Click on *Next*.

Create annotations on dependent tiers   ✕

Create annotations on dependent tiers

☑ phrase-segnum-en

Previous   **Next**   Finish   Cancel

2.7.3.3 If you want annotations to be created on every tier **tick all the boxes**. If you want to create the annotations on the **note tiers manually, do not tick those boxes**. Select **Empty Annotations** behind the ticked boxes. Click on *Finish*.

Create annotations on dependent tiers   ✕

Create annotations on dependent tiers

	Tiers	Empty Annotations	Annotation With Value of P.
☑	phrase-txt-arh	⦿	◯
☑	free_translation-gls-en	⦿	◯
☑	note-note-en	⦿	◯
☑	gloss_note-note-en	⦿	◯

☐ Overwrite the annotation values

Previous   Next   **Finish**   Cancel

Yellow empty annotations appear.

2.7.4 You can now **add text to the annotations** by **clicking twice** on an annotation. If you want to add **manual annotations on the note tiers**, **click twice** on the desired spot. A dependent annotation is added. Note that these tiers are dependent and the annotations, therefore, have the same length as their parent.
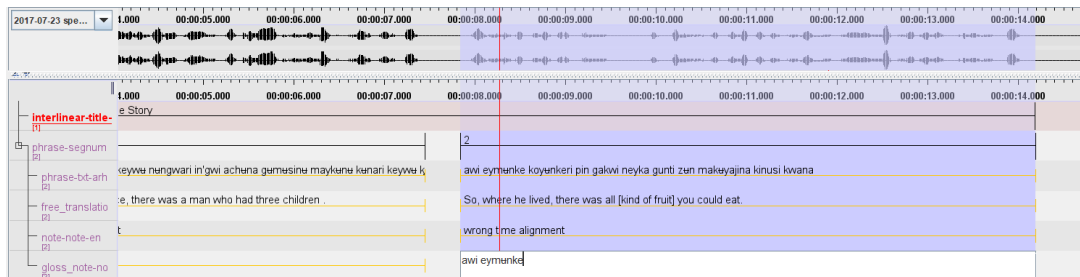
The finished result:
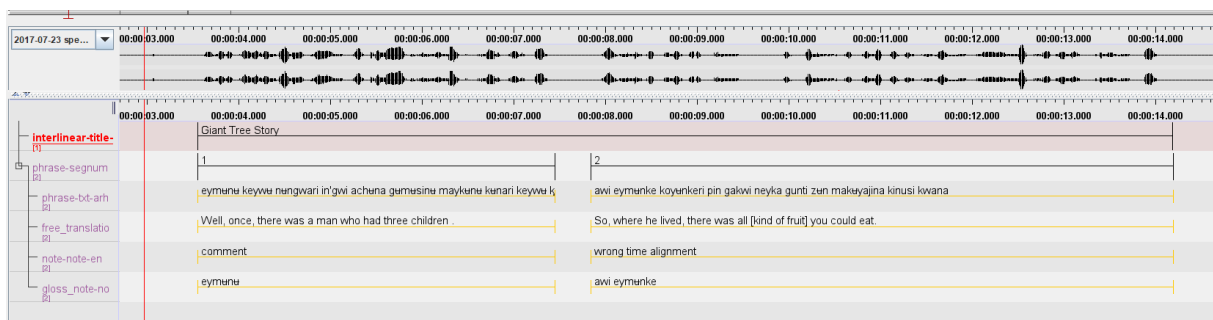
Your ELAN file is ready for export and import into FLEx.

3. Import ELAN to FLEx

3.1 Export ELAN

3.1.1 Click on *File > Export As > FLEx File*.
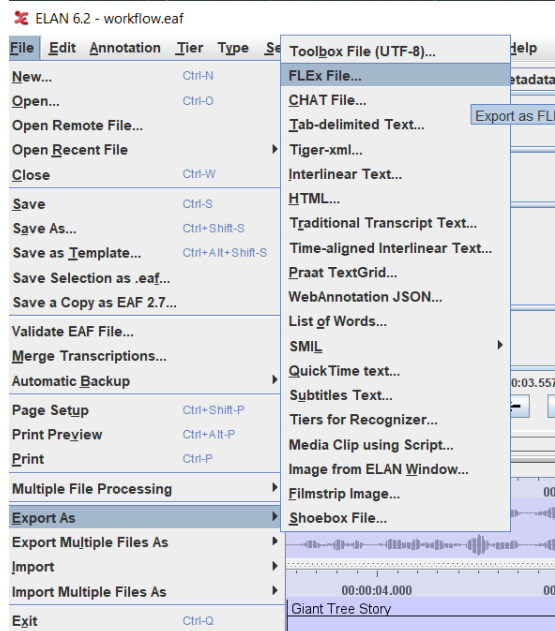
3.1.2 *Tick* the box *Export interlinear-text tier* and select *interlinear-title-en(g)*. Click on *Next*.[3]

---
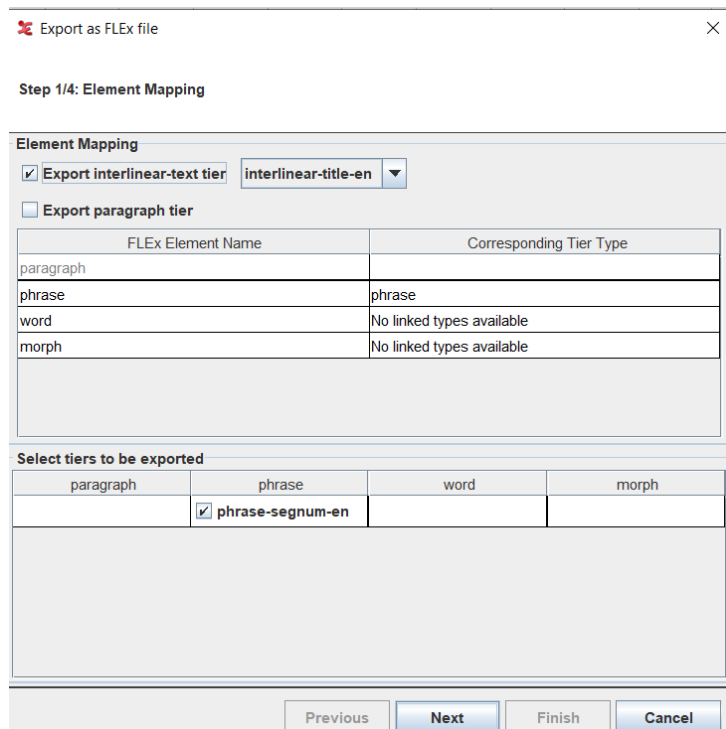
[3] If the tiers are set up incorrectly, it will be noticeable from this stage onward.

3.1.3 Click on *Next*. In case you do not want to export all the selected tiers, you can untick the boxes of the tiers that should not be exported. Elan offers the possibility to choose which tiers you want to export.

Export as FLEx file ✕

Step 2/4: Element-Item Configuration

☑ Select a tier type for 'morph-type' tiers | No linked types available ▼

Item Mapping

interlinear-text_item	phrase_item	word_item	morph_item
No linked types available	phrase-item	No linked types available	No linked types available
	text-is-translation		
	note		
	No linked types available		

Select tiers to be exported

interlinear-text_item	phrase_item	word_item	morph_item	morph-type
	☑ phrase-txt-arh			
	☑ free_translati...			
	☑ note-note-en			
	☑ gloss_note-n...			

Previous | Next | Finish | Cancel

3.1.4 Click on *Next*.

Export as FLEx file ✕

Step 3/4: Element-item 'type' and 'lang' attribute configuration

Specify the value for type and lang attribute based on

⦿ tier   ○ tier type

TierName	type	language
interlinear-text		
interlinear-title-en	title	en
phrase		
phrase-segnum-en	segnum	en
phrase-txt-arh	txt	arh
free_translation-gls-en	gls	en
note-note-en	note	en
gloss_note-note-en	note	en
word		
morph		

Type-Lang value configuration

Add/remove values for   ⦿ type ○ language

Add custom value   Add

Select the value to be removed   <select>   ▼   Remove

Previous | Next | Finish | Cancel

3.1.5 Click on **Browse** and **save** the file. Click on *Finish*.
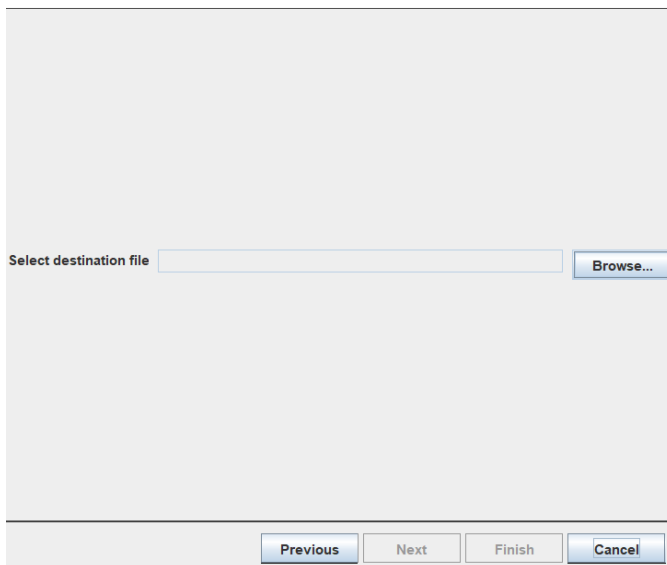
3.2 Import FLEx

3.2.1 Open FLEx. Go to **Texts and Words**. Then go to the tab **File > Import > Interlinear FLExtext**.
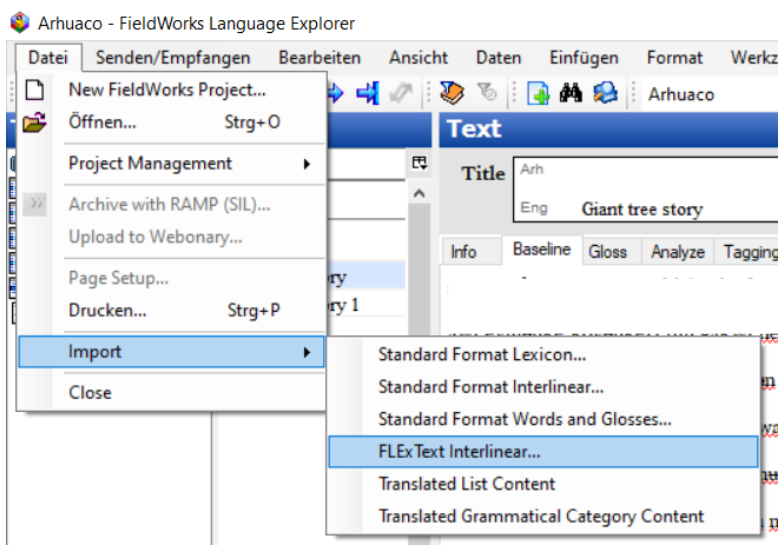
3.2.2 **Select a file** and click on *OK*.
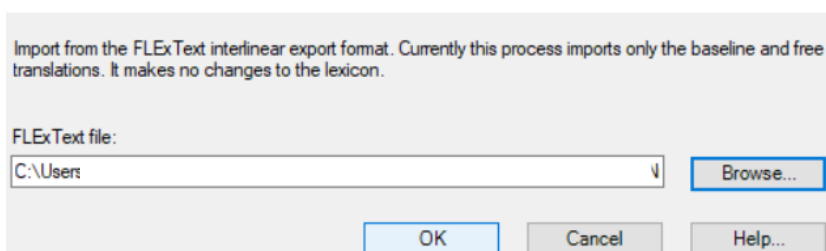
The result:
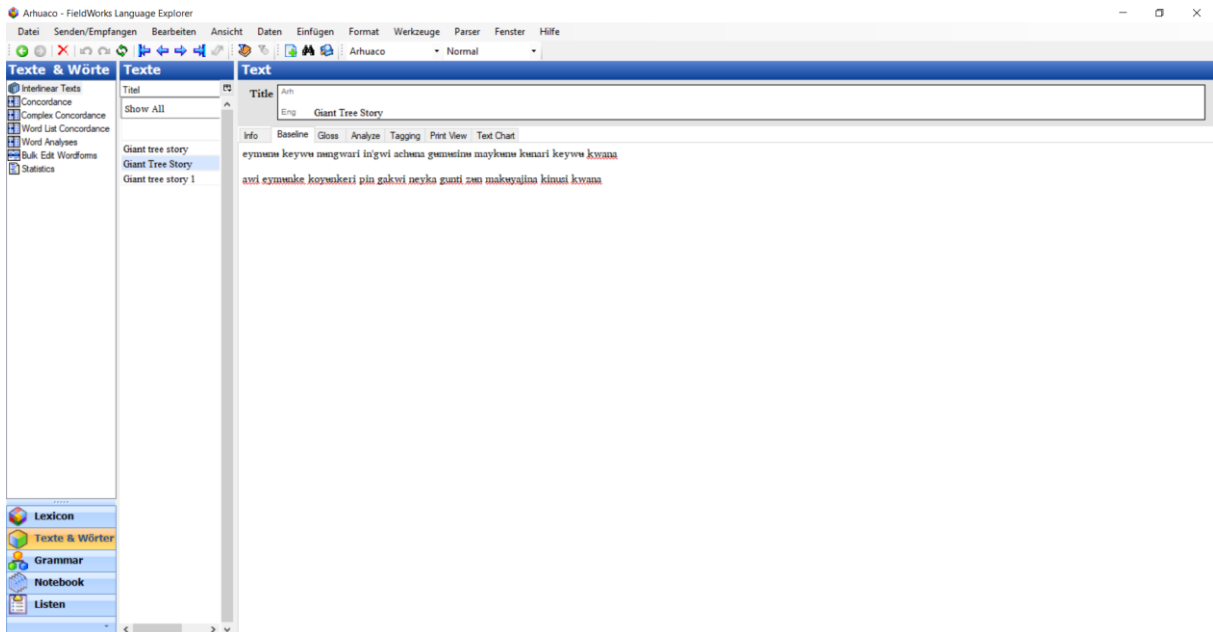
In the tab *Analyze* the free translation and the notes are visible. (Blue marks words already in the lexicon.)
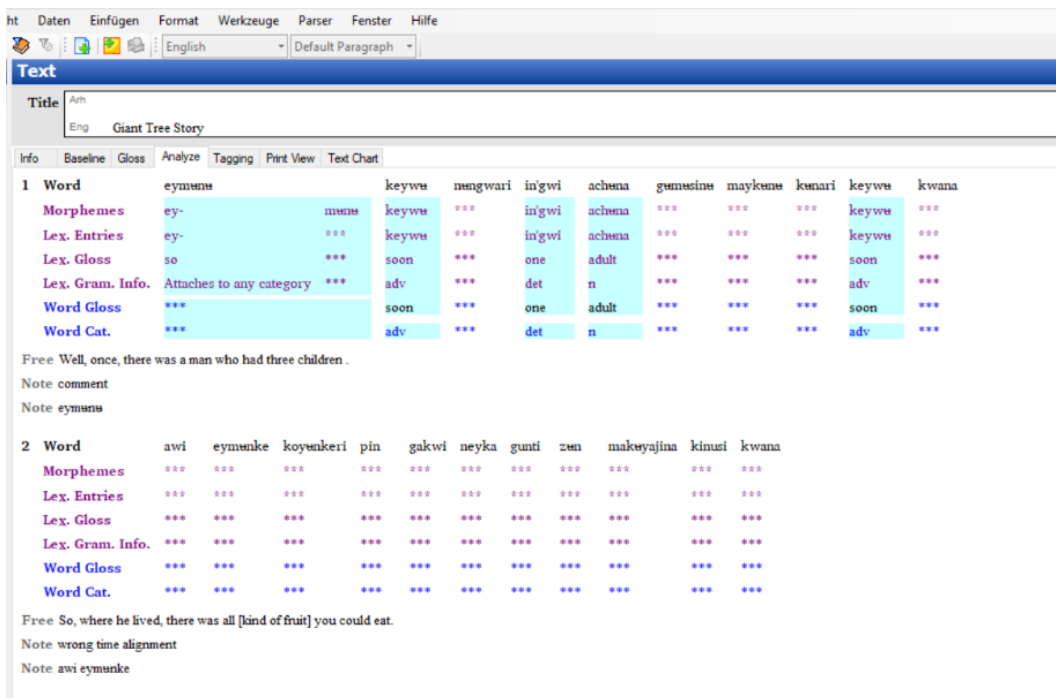
Note: Sometimes, the note tier is hidden; it does not appear automatically when opening an flextext-file. If this is the case, click somewhere below the sentence number so that a small arrow appears:

Then, click on the arrow, select *Add Line* and click on *Note*:

Then, the note tiers do appear:

4. Import FLEx to ELAN

4.1 Export FLEx

4.1.1 Go to *Texts and Words* and click on the tab *Analyze*.

4.1.2 Go to *File > Export Interlinear*.

4.1.3 Select *ELAN, SayMore, FLEx* and click on *Export*.

4.1.4 *Select the text* that you want to export. Click on OK
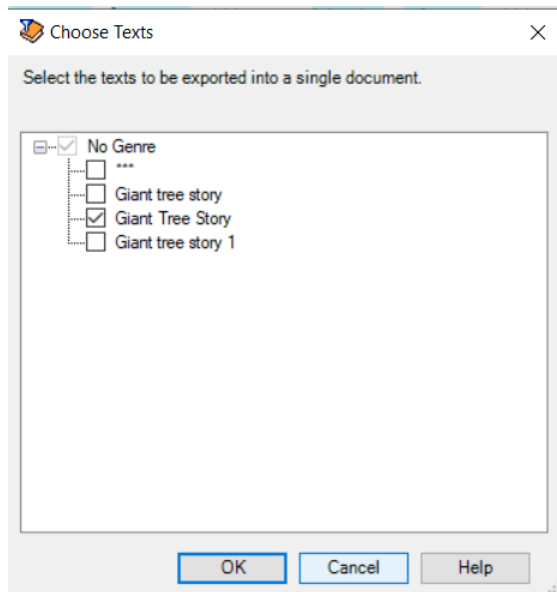
4.1.5 Save the file.

4.2 Import ELAN

4.2.1 Open ELAN. Go to *File > Import > FLEx File*.
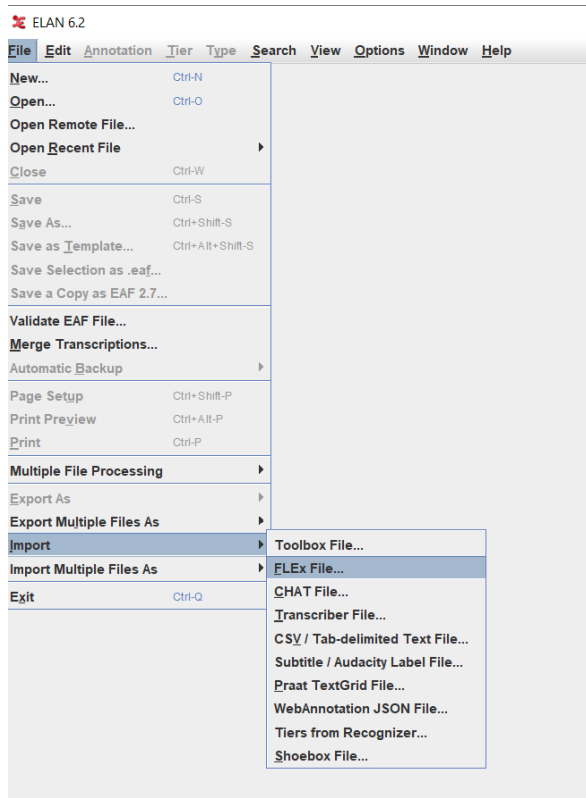
4.2.2 Click on the three dots and select a **FLEx file** and a **media file**. Click on *OK*.
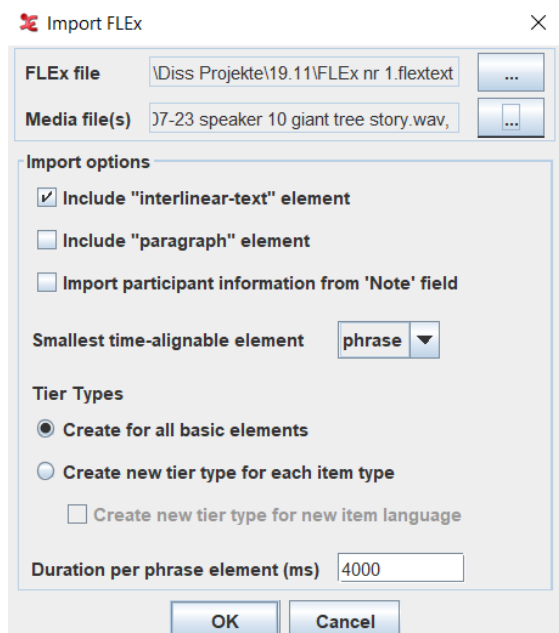
The result:
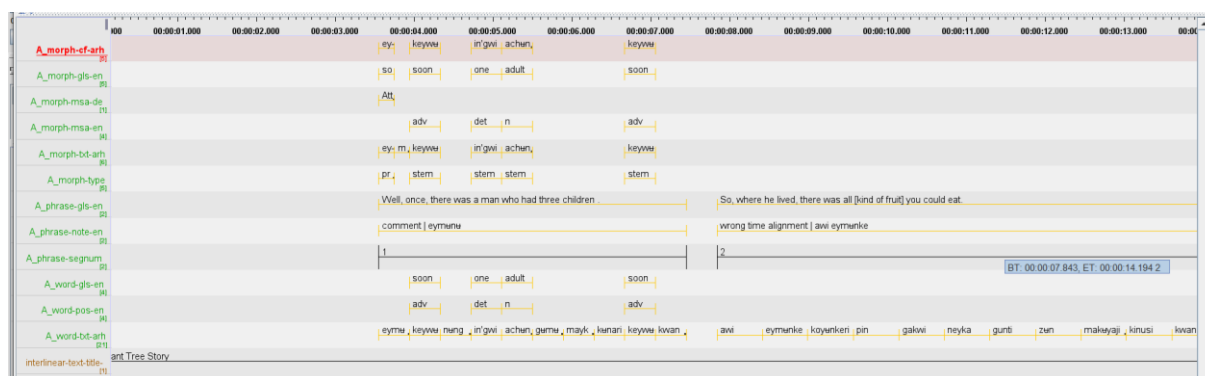
Note: When you import the FLEx file into ELAN the note tiers get merged. This means that the two note tiers that were created in the original ELAN file will automatically be combined into a single tier by FLEx.

5. ELAN-FLEx-ELAN with Multiple Speakers

5.1 Add a Second Set of Tiers to ELAN

5.1.1 Follow the steps described including section 2.5 with one small difference: To the *tier names*, *add "A_"*. The letter A stands for the first speaker. It is important that you separate the speaker name from the actual tier name with an underscore _. Otherwise, the import to FLEx will not work.

5.1.2 Then go to *Tier > Add new participant*
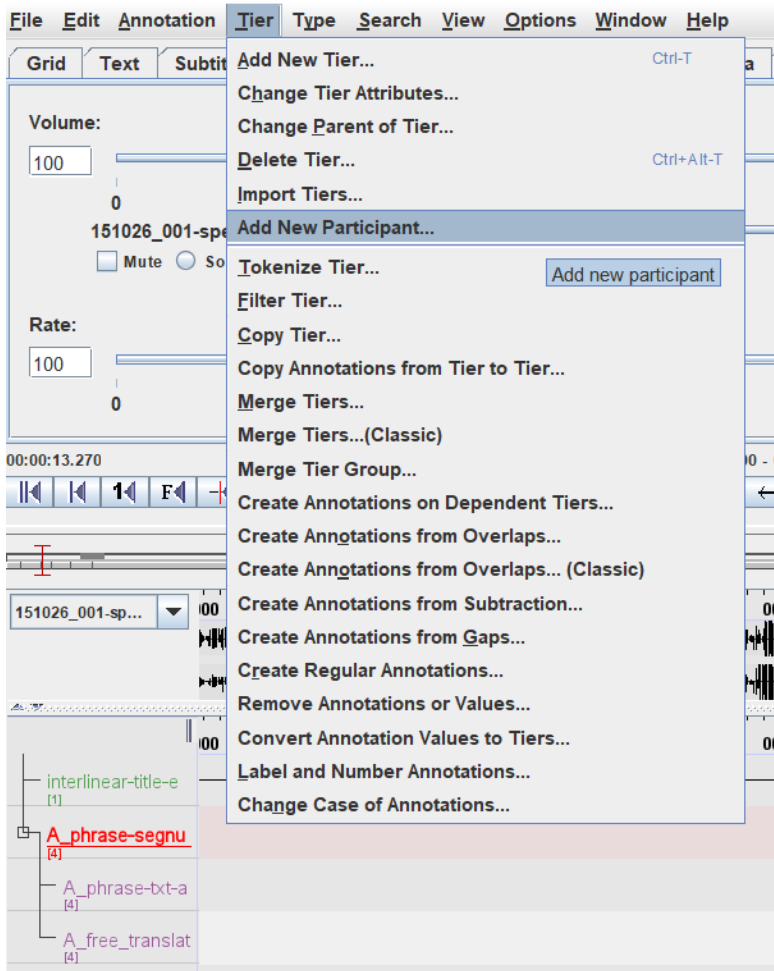
5.1.3 A window pops up:

- Select "A_phrase-segnum-en(g)"
- Specify the new participant. Type in "B"
- Enter the value to be replaced. Type in "A"
- Enter the new value for the replacement. Type in "B"
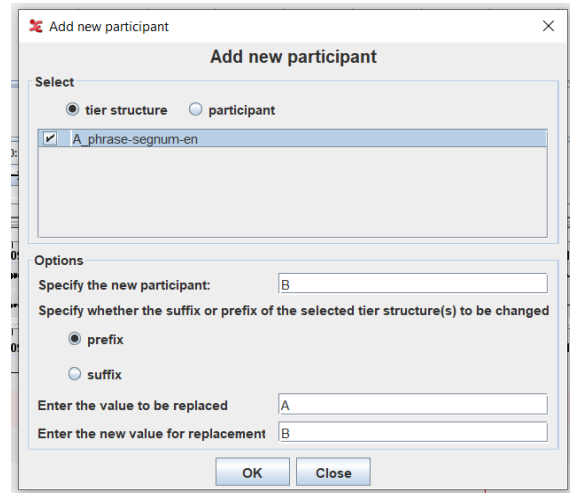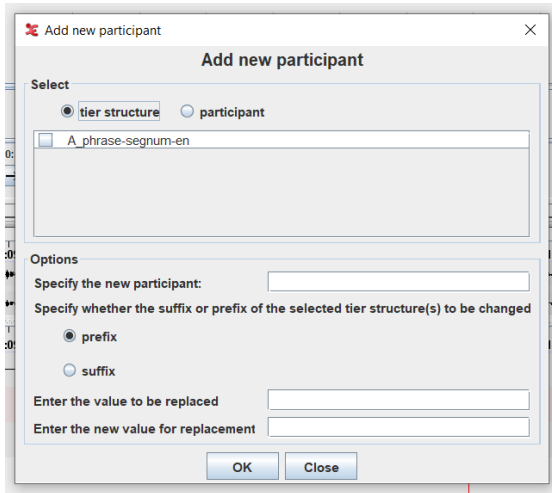
5.1.4 A new window pops up and the second set of tiers appears immediately. You can click on "close" on both windows.
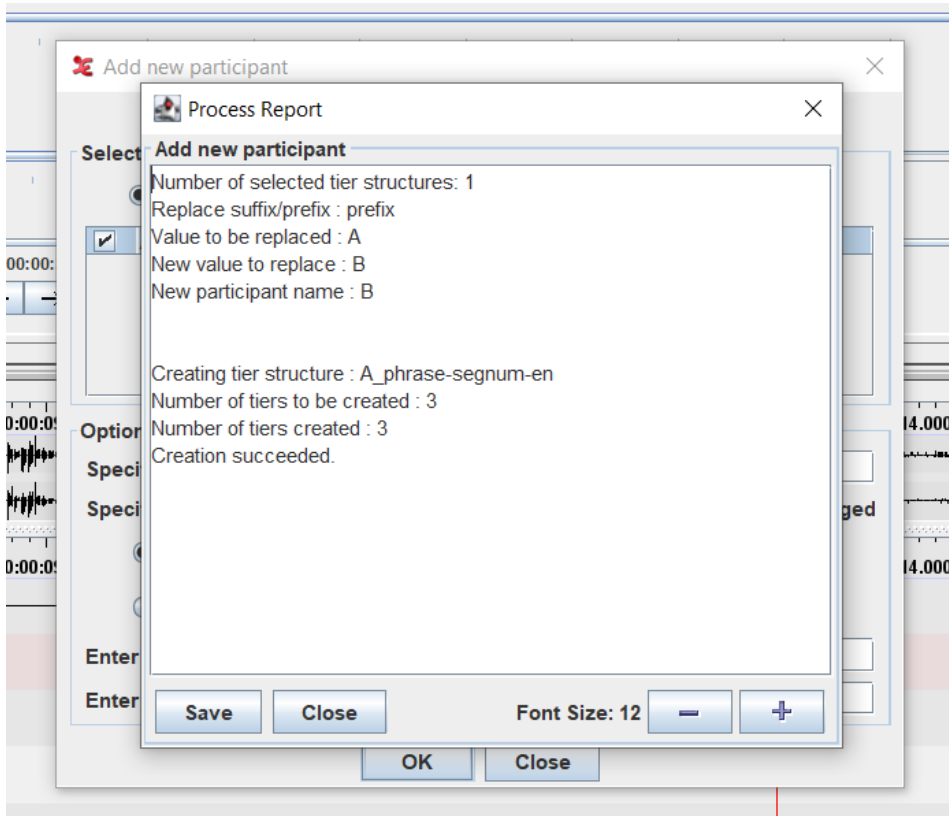
Now, you have the following tier structure:
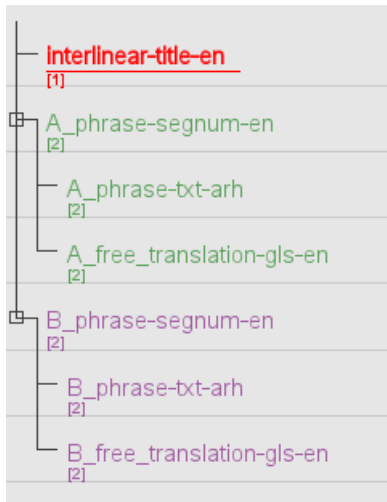
5.2 Annotation

You can now proceed with the annotation process as described above in section 2.6 above with the only difference that you have one set of tiers per speaker.

5.3 Export ELAN to FLEx

Again, follow the steps as described above in chapter 3. The difference is that you have to make sure that both set of tiers are selected during the process, which should be that way by default:

5.4 Import into FLEx

For import to FLEx, follow the steps described above. Here, the result in FLEx. Note that you do not know which speaker said what, unless you keep a list with sentence numbers and speakers.

5.5 Export FLEx to ELAN

Follow the steps exactly as described above in chapter 4. You do not have to change anything here. The result:

6. Supplementary Materials

The supplementary materials of this guideline consist of eaf (ELAN) and flextext files. The flextext files must be imported into FLEx and ELAN, as described in section 3.2 and 4.2. Before you can import the flextext ELAN-FLEx, you need to create a project with the corresponding vernacular language and the analysis language. After that, you can follow the instructions in step 3.2.

Section 2, 3, and 4 – General Workflow (Arhuaco – English)

PMT Arhuaco ELAN-FLEx.eaf	ELAN file created in ELAN
PMT Arhuaco ELAN-FLEx.flextext	Exported ELAN file that is ready for import into FLEx
PMT Arhuaco FLEx-ELAN.flextext	Exported FLEx file that is ready for import into ELAN
PMT Arhuaco FLEx-ELAN.eaf	The complete (glossed) data from FLEx in ELAN (result after import)

Section 5 – Multiple Speakers (Arhuaco – English)

PMT Arhuaco ELAN-FLEx two speakers.eaf	ELAN file created in ELAN
PMT Arhuaco ELAN-FLEx two speakers.flextext	Exported ELAN file that is ready for import into FLEx
PMT Arhuaco FLEx-ELAN two speaker.flextext	Exported FLEx file that is ready for import into ELAN
PMT Arhuaco FLEx-ELAN two speakers.flextext	The complete (glossed) data from FLEx in ELAN (result after import)

Example – Bats (Bats – English)

PMT Bats ELAN-FLEx.eaf	ELAN file created in ELAN
PMT Bats ELAN-FLEx.flextext	Exported ELAN file that is ready for import into FLEx
PMT Bats FLEx-ELAN.flextext	Exported FLEx file that is ready for import into ELAN
PMT Bats FLEx-ELAN.eaf	The complete (glossed) data from FLEx in ELAN (result after import)

7. References

Gaved, Tim; Salffner, Sophie (2014). *Working with ELAN and FLEx together: an ELAN-FLEx-ELAN teaching set.* The Language Archive: Third Party Resources. https://archive.mpi.nl/tla/elan/thirdparty (Accessed: 2021-09-16).

Hellwig, Birgit et al. (2021). *ELAN – Linguistic Annotator (Manual).* Max Planck Institute for Psycholinguistics Nijmegen. https://www.mpi.nl/corpus/html/elan/index.html. (Accessed: 2021-11-30).

Max Planck Institute for Psycholinguistics Nijmegen. (2021). *ELAN.* The Language Archive: Download. https://archive.mpi.nl/tla/elan (Accessed: 2021-06-01).

SIL International. (2021). Fieldworks Language Explorer (FLEx). https://software.sil.org/fieldworks/download/ (Accessed: 2021-06-01).