

Introduction to ELAN (with Audacity) for language transcription

Last revised August 14, 2026

Opening comments

Downloading our transcription tools

- Resource 1: Software downloading directions

Getting to know ELAN

- Downloading our example files for this session
- Resource 2: Setting up an ELAN file
- Resource 3: Segmentation mode, copying to tiers, and transcription mode
- Activity: creating annotations with ELAN

Getting to know Audacity

- Downloading our example file for this session
- Resource 4: Audacity

Mobilizing transcriptions

- Resource 5: Data management
- Downloading our example files for this session
- Resource 6: Exporting to text
- Resource 7: Corpus searches
- Activity: searching for keywords in a narrative

(These resources are adapted from materials prepared in earlier workshops with McGill University, Kanien'kéha:ka Onwawén:na Raotitióhkwa Language and Cultural Center, and Tsi Ronterihwanónhnha ne Kanien'kéha.)

Opening comments

This packet is intended for a general introduction to ELAN, as well as Audacity, for purposes of language transcription. It is an adaptation of materials previously used in contexts of transcribing specific languages, namely Kanien'kéha and Ktunaxa. In creating a general resource, we do not include links to transcriptions or recordings in the hands-on activities.

Resource 1: Software Downloading Directions

ELAN & Audacity

ELAN 7.1

All transcriptions will be done using ELAN software, which allows users to create time-aligned written transcriptions of audio or video files, and create searchable annotations of audio or video files. This can be useful to create printable transcripts and video subtitles, for instance. The time-alignment function provides a significant advantage over transcription using plain text tools like Word, because transcriptions are associated with specific sections of the media file, allowing for the creation of subtitles that match the audio and video. ELAN files can be exported into a variety of formats, including plain text, and ELAN also offers advanced search functions, allowing users to search for very specific contexts of use of a particular word, which should be very useful to teachers, learners, and researchers alike. It is the standard software used by most language researchers to create transcriptions for language documentation, and is fairly easy to learn and use with significant documentation and online support.

To download ELAN 7.1, click [here](#) to find multiple versions that are free for your specific OS platform (Mac, Windows, Linux, others). Make sure you download the full version of ELAN, not versions whose name begins with Simple.

Audacity

Audacity is audio editing software, which allows users to manipulate audio files in order to prepare them for transcription in the ELAN software (e.g., modify loudness, remove background noise, cut excerpts from a recording, etc.). It is especially useful for trimming and cutting audio files, for instance if transcribers wish to work on only a fragment of a radio show episode, or if only a portion of an episode qualifies as part of the “green” level while the rest is more sensitive.

To download Audacity, click [here](#) to find multiple versions that are free for your specific OS platform (Mac, Windows, & Linux).

Learn more about ELAN: <https://archive.mpi.nl/tla/elan>

ELAN online manual: https://www.mpi.nl/corpus/manuals/manual-elan_ug.pdf

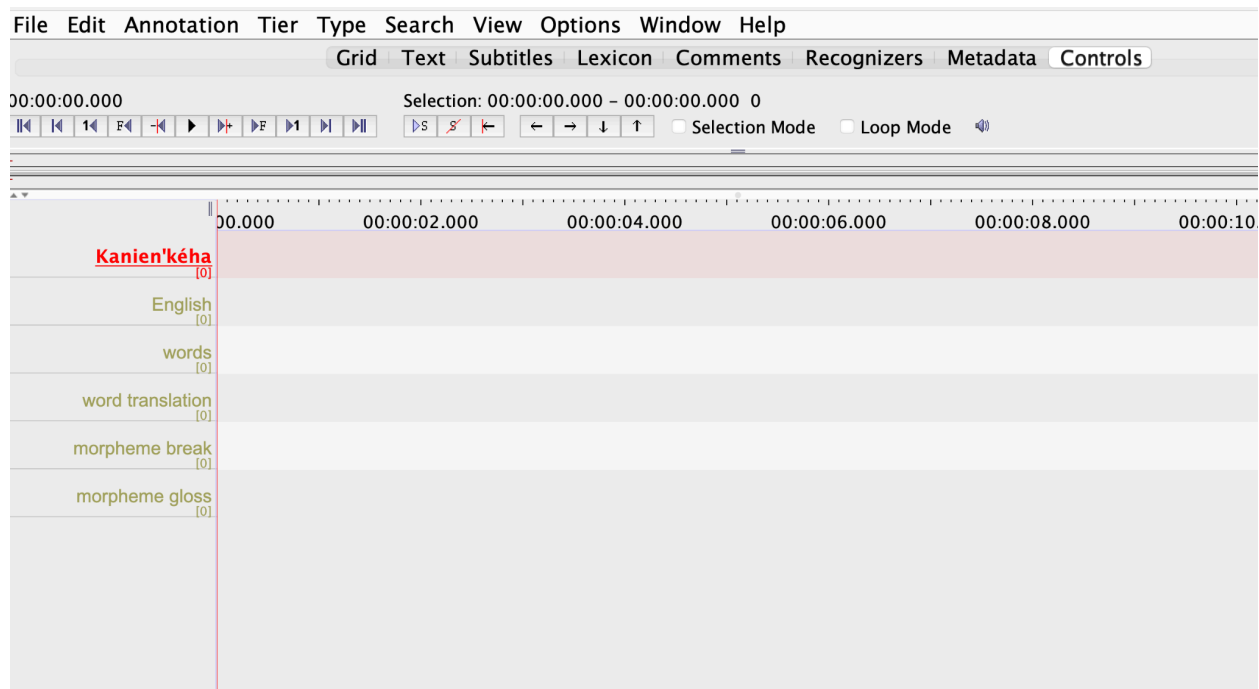
Learn more about Audacity: <https://www.audacityteam.org/FAQ/>

Audacity online manual: <https://manual.audacityteam.org/>

Getting to know ELAN

Downloading our example files

As with any transcription, we first need the media itself. In this case, we will use an .mp4 video file. We will also use a .etf template file, which automatically loads a basic structure for our transcription in ELAN. With a template file, you can save time by immediately having your desired tiers in place, as shown below.



1. Identify where you would like to store this transcription on your computer: for this example, let's create a new folder called “ELAN tutorial” in the “Desktop” directory.
2. Move/copy at least one media file into your folder; this media file should be either an .mp4 video, .mp3 audio, or .wav audio.
3. If you have an ELAN template file (file extension .etf), move/copy that file into your folder.

We will use our new “ELAN tutorial” folder throughout today’s workshop, starting with Resource 2: Setting up a file.

Resource 2: Setting up a file

How to start a new transcription file

1. Open ELAN and start a new transcription: **File > New...** (or ⌘N)
2. Select your media: **Add Media File...**
3. Select your template file: **Add Template File...**
4. Click **OK**

How to add another linked file

(Adding another linked file can be helpful when, for instance, you'd like to add a .wav audio file so that you can see the waveform in ELAN alongside the .mp4 video.)

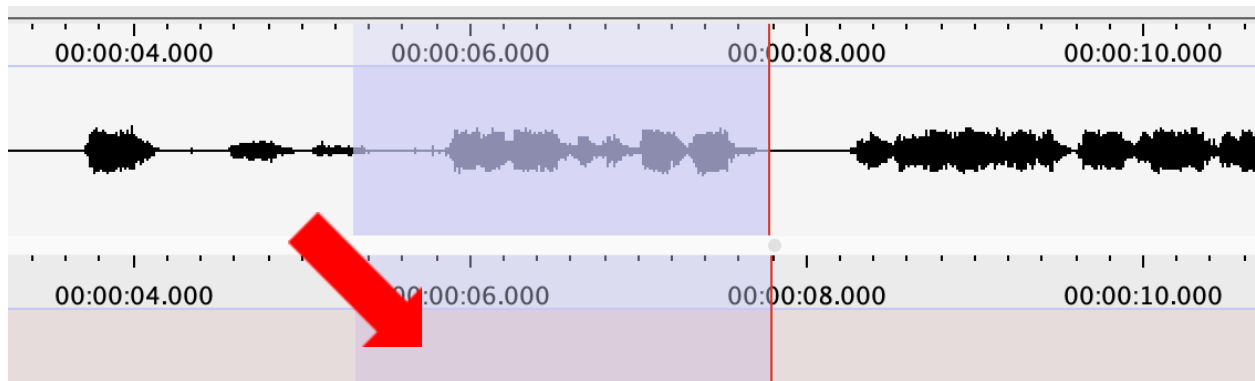
1. Go to **Edit>Linked Files...**
2. Select **Add...** and locate the additional media file, i.e., a media file that is distinct from the media file you added when you started the transcription file.
 - a. Make sure that all linked media files share the same audio. Otherwise, they will be played out of sync.
3. Select **Apply** and then **Close** to close the Linked Files window.

How to set up auto-save

1. Save your file manually: **File> Save** (or ⌘S)
2. Set up your automatic backup: **File > Automatic Backup > 1 minute**

How to make an annotation on a tier

1. Highlight the portion you would like to annotate on the waveform by sliding your cursor across it.

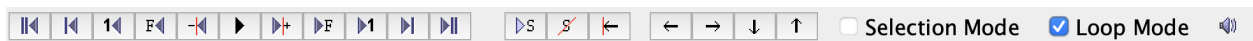


2. Double-click on the tier you want the annotation to be on, identified by the arrow in the image above. The selection box will then turn white with a text cursor (the blinking vertical line).
3. If you would like to add text to the annotation, you can type the annotation directly and then press Enter to complete the annotation. If you want to create an empty annotation, simply press Enter without adding text.

Note: The first annotation needs to be on a parent tier (the topmost tier on your template).

How to make the audio selection repeat

1. Highlight a portion of the audio, or select an annotation
2. Check the **Loop Mode** box (or ⌘L), and use the **►S** button to play the selected portion



► plays from your pointer, ►S plays selected part Loop mode

How to slow down the speed of the recording

1. On the *Controls* view, find the *Rate* slider, **slide the cursor left** to slow down the speed of the recording; **slide the cursor right** to speed up the recording.



How to make sure the tiers appear in the right order

1. You can **drag and drop** tiers on the left-side list to re-order
2. To make the tiers appear based on their relationship (to have all the tiers related to a speaker show up together, for instance), **right-click** on the left-hand side of the ELAN screen where the tiers are listed. A menu box will appear. **Select sort tiers**, and then **sort by hierarchy**.

Useful keyboard shortcuts

- To start and stop the media file while you are transcribing
 - Select an annotation. If transcribing, double-click on the annotation so that it turns white

- Press **Shift + space bar**. The recording starts playing.
- Press the same two keys to stop the recording.
- To activate or deactivate loop mode
 - Press **Command (⌘) + L** once to activate and once to deactivate
- To enter an annotation

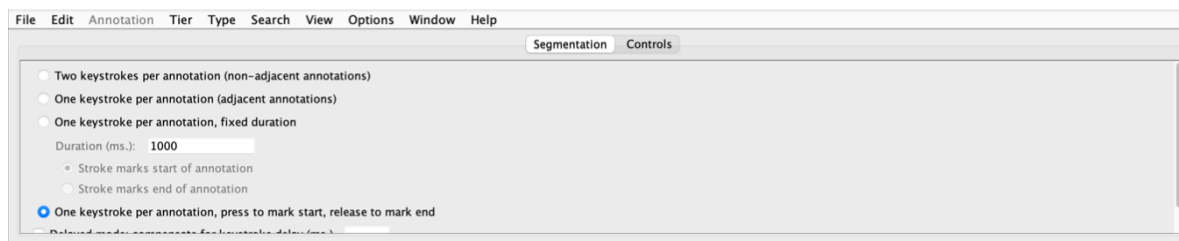
Press **Enter** when you are done transcribing

Resource 3: Segmentation mode, copying to tiers, & transcription mode

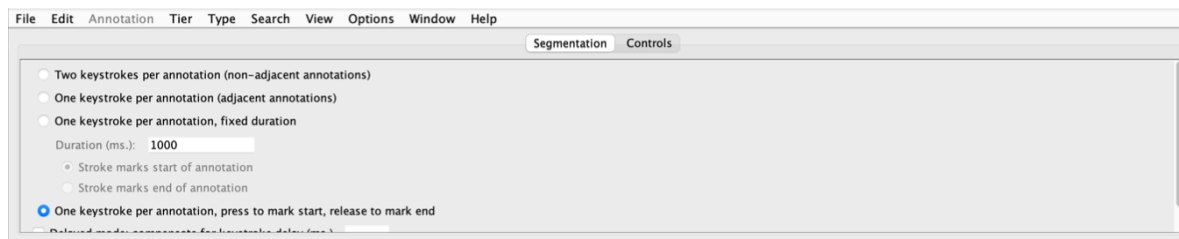
Part 1: Segmentation mode

Segmentation mode allows you to segment the media file you are transcribing into smaller annotations, for instance as single utterances or sentences, using only keystrokes. This means that you can segment an entire tier for a given speaker before transcribing. Segmentation mode does not allow you to transcribe, so once you are done segmenting, you must change to annotation mode or transcription mode (through the *Options* menu in the top menu bar.)

1. To use segmentation mode go to **Options > Segmentation Mode**
2. Depending on the recording that you have, it will determine which segmentation mode you will like to use.
 - Two keystrokes per annotation (useful for recordings where sentences are separated by pauses, and for recordings with more than one speaker)
 - One keystroke per annotation (useful for recordings where a single speaker speaks continuously)
 - One keystroke per annotation: press to mark start, release to mark end



3. You will need to choose the speaker on the left side and it will highlight in **RED**.
4. Once you press play the first time you can use shortcut keys below to control when your annotation will begin/stop.
5. Controls for speed can be found in controls under rate



Shortcut keys

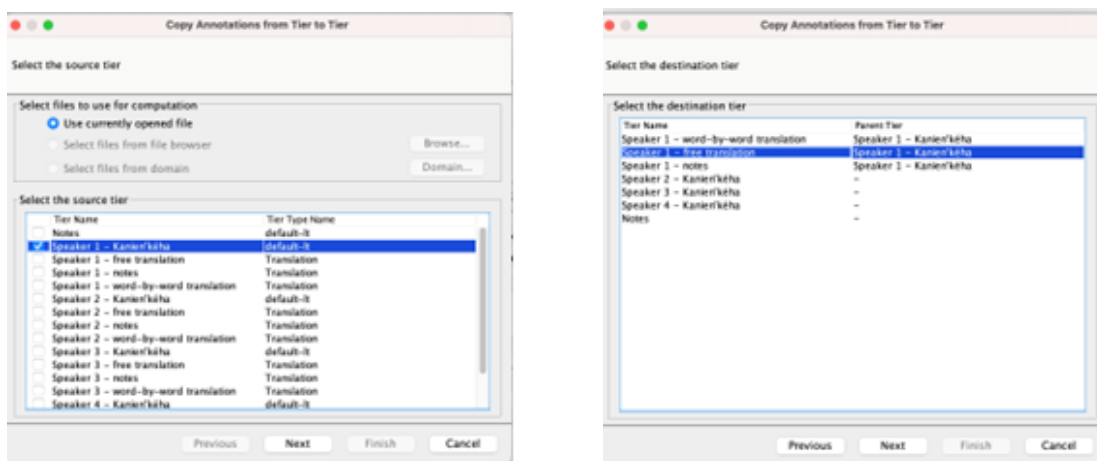
- “Space bar” – plays the segment
- “Enter” – marks the beginning/ending of segment

Part 2: Copy an annotation from tier to tier

(Note: this Part 2 might no longer be needed, given new updates in ELAN's version 7.1.)

Once you have segmented a speaker tier, you might want to re-create the same annotations on a dependent tier. This is particularly useful if you are using transcription mode (described below).

1. Go to **Tier > Copy Annotation from Tier to Tier**.
2. Choose your source tier (the tier where annotations will be copied from). Click **Next**.
3. Choose the destination tier (the tier where annotations will be copied to).
4. Set to **All Annotations** and press **Finish**.



Part 3: Transcription mode

Transcription mode allows you to transcribe pre-segmented annotations in a table-like format, as shown in figure 1.

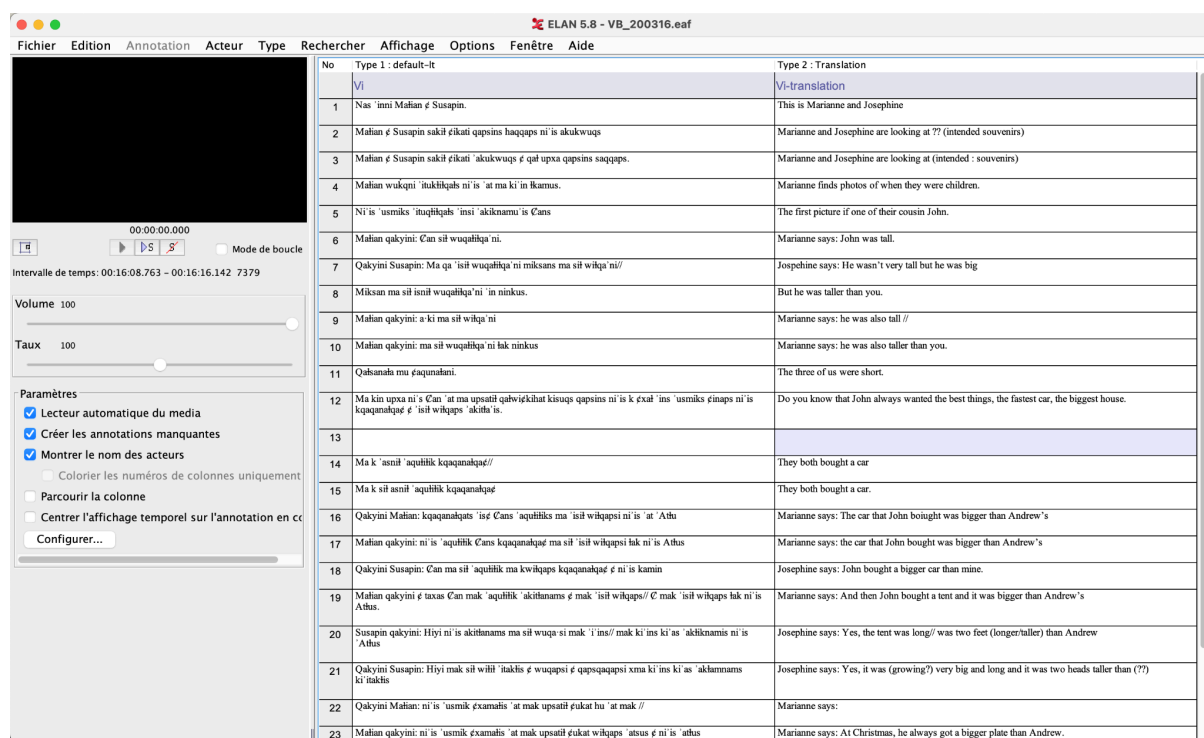


Figure 1: ELAN's transcription mode, showcasing a transcription in Ktunaxa alongside its translation. The speaker is Violet Birdstone.

1. If it's the first time you enter transcription mode for this specific file, you will be prompted to choose which tiers you want to display. You can change the display at any point after this by clicking on [configure](#) (red arrow in Figure 2).
2. Note that ELAN can create missing annotations on dependent tiers. To ensure that this is done, make sure that the box [Create missing annotations](#) (blue arrow in Figure 2) is ticked when you configure your transcription mode settings.

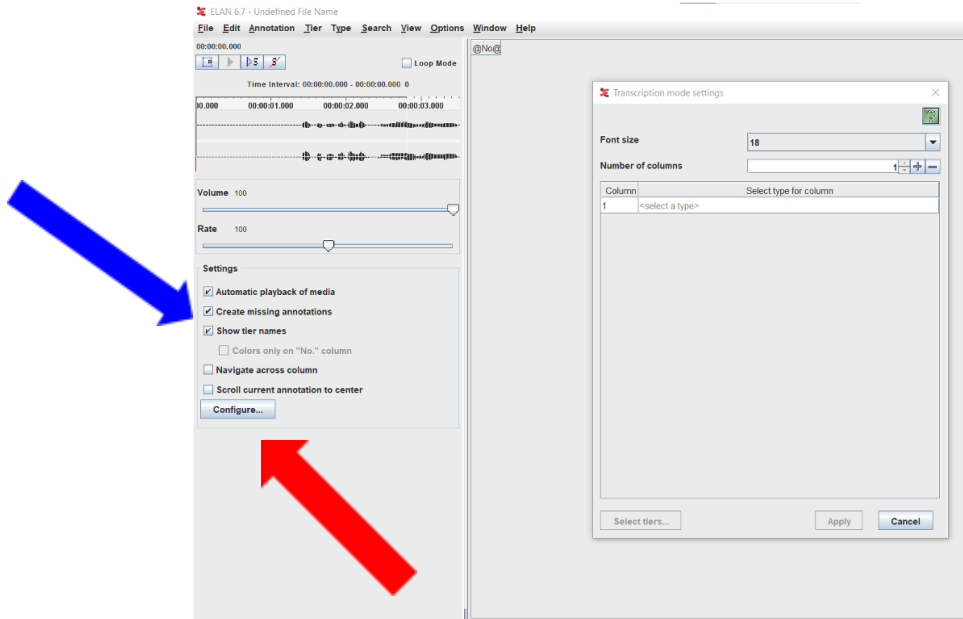


Figure 2: Transcription mode in ELAN

- Before you choose which tiers are represented in your table, you need to tell ELAN which types of tiers you wish to represent. If you want to have a speaker's speech alongside its translation, you should choose *default* and *translation*, as shown in Figure 3.

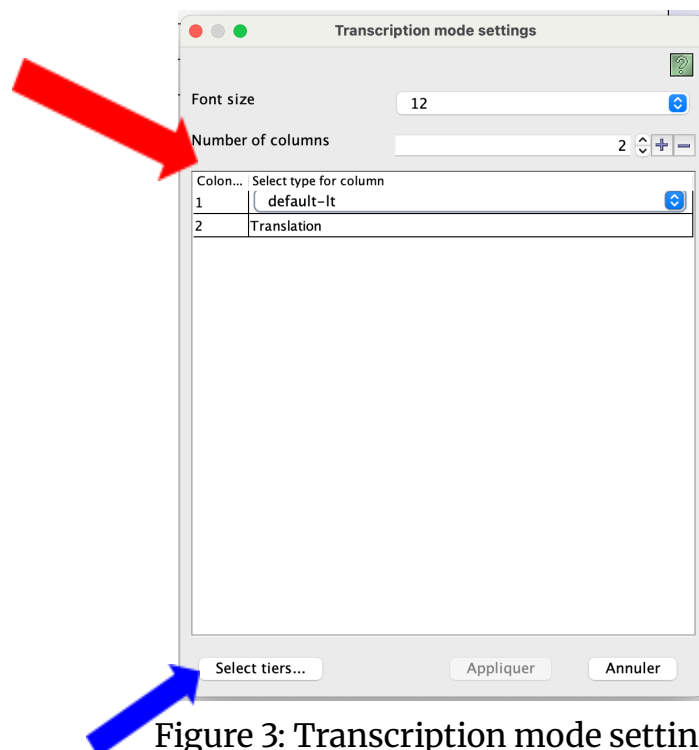


Figure 3: Transcription mode settings

4. Next, you have to tell ELAN which set of tiers you want to display. Click on
- *Select tiers* (the blue arrow in Figure 3 above). A new box will open. You can choose to display a single speaker and its tiers, or you can display multiple speakers and their tiers. Check the box(es) for the speaker(s) you want to display, and click apply (in Figure 4 below, *appliquer*)

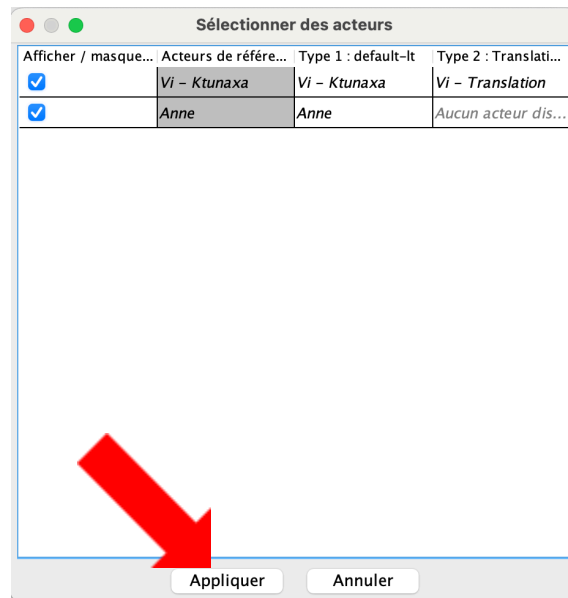


Figure 4: Select tiers dialog box

5. To use transcription mode, you must have segmented your recording into annotations, either in annotation mode or in segmentation mode. You cannot adjust annotation boundaries, merge, or split annotations in transcription mode, but it is easy to go back and forth between annotation mode and transcription mode to make minor adjustments to the length and location or sound alignment of your annotations.
- For instance, you might have cut off a word at the beginning or end of an annotation, you might want to merge two annotations that are related, or split an annotation in two because it is too large.
 - This will have to be done in annotation mode.
6. To access transcription mode, go to
- *Option > Transcription mode* or
 - Right-click on the annotation you want to modify, a drop-down menu opens, select *Edit in annotation mode*.

Activity: creating transcriptions with ELAN

Getting familiar with transcription controls in ELAN

Using Resource 2 as a guide—alongside help from your peers or the facilitators—try to accomplish each of the following in your example ELAN file.

1. Set up auto-save for your .eaf file.
2. Create one annotation on the parent tier: for instance, a transcription of a phrase or utterance in the language that you are documenting.
3. Create one annotation on a child tier: for instance a translation of the phrase or utterance you transcribed in step 2.
4. Make an audio selection run on a loop.
5. Adjust the speed of the audio playback.
6. Change the order of the tiers.
7. Challenge 1: Edit an annotation that you've already created.
 - a. (Hint: for Windows, try right-clicking on the annotation; for Mac, try Control+clicking on the annotation.)
8. Challenge 2: Delete an annotation that you've already created.

Practicing with segmentation and transcription modes

1. Follow Part 1 of Resource 3 to create segmentations on the utterance tier for the first ~30 seconds of the media file. You can use pauses as a cue for when to start and end segments. (Remember that the segments intentionally start off empty. We'll add the written transcriptions soon!)
2. Follow Part 2 of Resource 3 to copy the empty segmentations onto the "Free translation" tier.
3. Follow Part 3 of Resource 3 to fill out the segments in the language that you are documenting (left column) and English (right column).

Getting to know Audacity

Downloading our example file

As Audacity is used for editing sound files, please be sure to download an audio file (file extension .mp3 or .wav) if you did not already download one for the earlier ELAN activities.

1. Save the audio file to a folder called “ELAN tutorial” in the “Desktop” directory.

We will use the contents of the “ELAN tutorial” folder in Activity 4: Audacity.

Resource 4: Audacity

1. Opening a sound file in Audacity

- Make sure that the sound file is one of the formats that are accepted in Audacity: mp3, m4a, AIFF, FLAC, WAV, and others.
- Navigate to *File > Open*. A dialog box opens, and you can then choose the file you want to work on.

2. Looking at the sound wave

- Stereo recordings will have two waveforms (two tracks). If sound quality is not a concern, you can merge these waveforms to create a mono recording (with a single track), but you don't have to.
 - In the top menu, select *Tracks > Mix stereo down to mono*.
- Depending on the length of your recording and your needs, you might want to zoom in and out to focus on specific areas of the recording.
 - In the top menu, select *View > Zoom in/out/etc.*
 - You can also use keyboard shortcuts as they are given in the *View* menu.
 - Zoom in: *cmd+1* or *ctrl+1*
 - Zoom out: *cmd+3* or *ctrl+3*

3. Listening to the recording

- Playback commands are clickable icons on the top left corner of the audacity window.



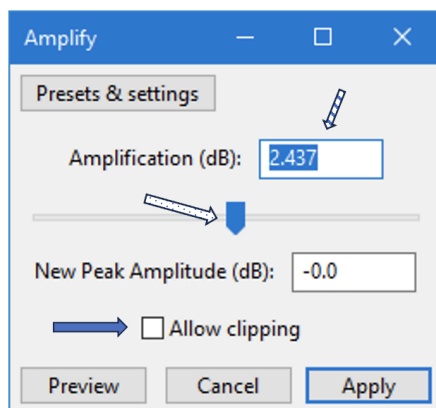
- Depending on your Audacity default settings, you might have to play around a bit to get the sound to come out through the right device.
 - In the top menu, select *View > Toolbars > Check Device toolbar*.
 - In the device toolbar, which should appear the playback toolbar on the top left corner, click on the sound icon.



- A dropdown menu appears, select the right device.
 - If a device that is otherwise properly plugged in does not appear in the dropdown menu, go to *Transport > Rescan audio devices*.

4. Cleaning up a sound file

- a. Adjusting the general volume (if you have a very quiet speaker or a poor-quality recording).
 - i. In the top menu *Effect > Amplify* OR *Effect > Volume and compression > Amplify*



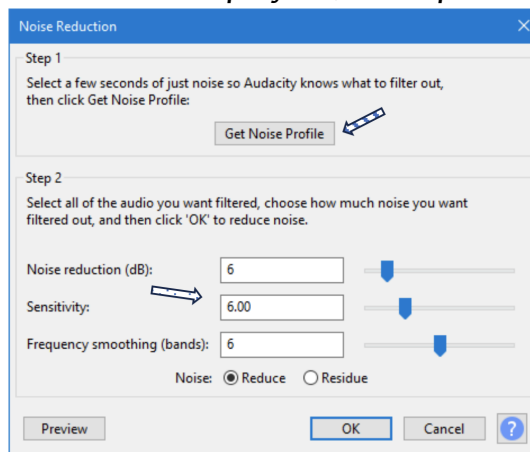
In the box above, you'll have to input an amplification value (the striped arrow). This value corresponds to the number of decibels you want to add. Decibels are the unit of measure used for volume. You can also use the slider (the dotted arrow) to determine the amplification value.

If the amplification value is so high that it will induce clipping, you won't be able to click on *apply*, unless you have checked *allow clipping* (the solid arrow).

Clipping happens when the amplitude of a recording exceeds what the sampling rate of your recording supports, such that the sound wave will be literally clipped above the maximal amplitude allowed (see below for a representation of clipping). For conversational recordings where you are not performing detailed sound analysis, you shouldn't worry too much about clipping.

- b. Removing background noise (an electric hum, an air conditioner, a fan, cars passing without interruption).
 - i. First, you'll need to teach Audacity to recognize the background noise you want to remove.
 - In your recording, select a small portion where there is only background noise (no overlapping conversation).

- In the top menu, *Effect > Noise reduction*.
Or *Effect > Noise removal and repair > Noise reduction*
- Click *Get noise profile* (the striped arrow).



- The portion of your file where you want to remove the background noise (possibly the entire file).
- In the top menu, *Effect > Noise reduction* Or *Effect > Noise removal an repair > Noise reduction*.
- Play around with the values in *Noise reduction* and *sensitivity* (dotted arrow) until your recording comes out clear enough. Leaving the *frequency smoothing (bands)* value to 0 ensures that the speakers' voices will remain as clear as possible.

5. Cutting up a file

- There are many reasons why you would want to split or remove parts of a recording: you might have long periods of silence that do not need to be noted in the transcription, long interruptions, several conversations or stories that would be better transcribed separately, sensitive information that you simply do not want to transcribe, etc.
- Removing a part of a recording.
 - Select the part of the recording you want to remove.
 - In the top menu, *Edit > Cut*.
- Splitting a recording into several files.
 - The simplest way to split a recording into several files is to cut and paste parts of a recording into new Audacity files.
 - Select the portion of the recording you wish to turn into an independent file.
 - In the top menu, go to *Edit > Cut/Copy*.
 - Open a new Audacity file: *File > New*.
 - Paste the cut or copied section of the original sound file, *Edit> Paste*.

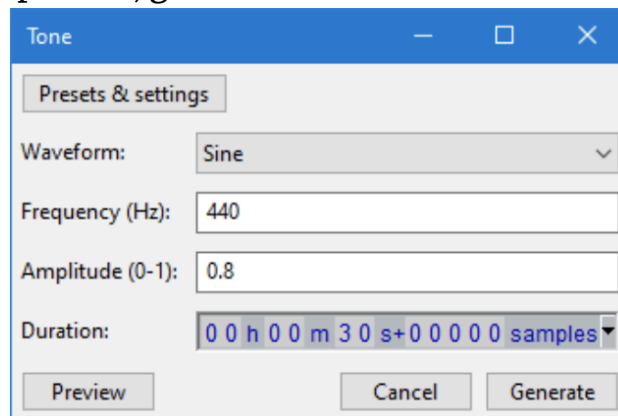
- vi. If you have used a *copy and paste strategy*, you might want to go back to your original file to remove the copied section.

6. Merging two files

- a. There are many ways to do this. The simplest one might be to *copy and paste* a file at the end of another file, using the commands in the *Edit* drop-down menu.

7. Covering sensitive information with a beep/white noise

- a. Select the portion of the soundwave you wish to cover.
- b. In the top menu, go to *Generate > Tone*

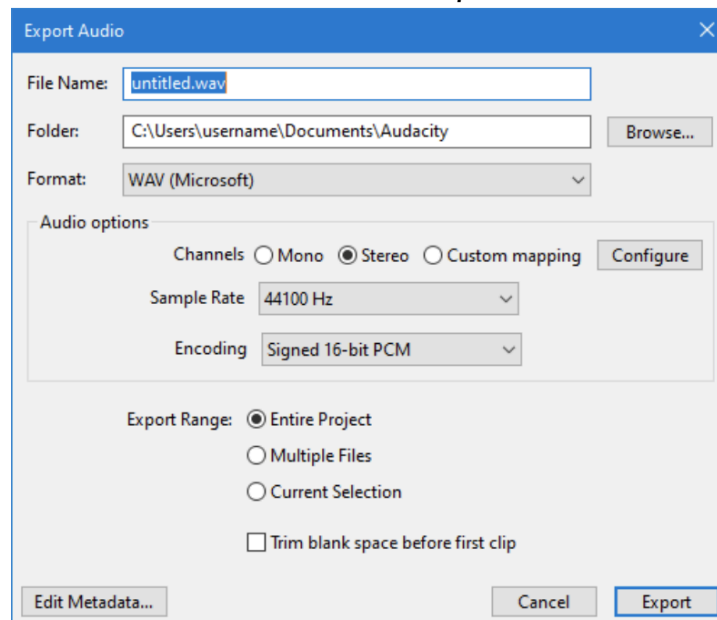


- c. In the dialog box, you'll have to input values for the frequency. I suggest using 220 which is a relatively low pitch A (the note), and a low amplitude of 0.1 or even lower, to ensure that the tone isn't unpleasant to listen to.

8. Saving/exporting

- a. Audacity has its own file format, which allows you to save a project and pick it up later. This is useful if you have a lot of work to do on a recording and you can't complete it in a single session. To save a project, go to *File > Save project > Save project as*.
- b. If you have done all the work you needed to do, you can export your file into a wave format. This is the file you'll be using as your media

file with ELAN. Go to *File > Export audio*.



- c. You have the possibility of specifying *audio options*, but we recommend you leave them as they are.
- d. In the Export Audio window, you can select **Edit Metadata...** to add custom labels for fields such as speaker name(s), topic, and date.

Resource 5: Data management

Metadata

When creating a transcription, it is important to keep track of *metadata*, which is an umbrella term for descriptive information that accompanies the transcription itself. Metadata includes, for example:

- The name(s) of the speaker(s)
- The date and place where the recording was made
- The names of the people involved in recording and transcribing
- The language(s) and dialect(s) represented in the recording
- The topic(s) discussed in the recording

There is no one-size-fits-all approach to metadata: different projects might call different kinds of metadata. One principle holds in any project, though: it is best to be **proactive** in defining what kinds of metadata you want to include before distributing your transcriptions. That way, future readers or listeners who access your transcriptions will not have to guess as to where this data came from.

ELAN has built-in tools for storing metadata, and here we focus on a couple of categories. Go to “Tier” > “Change Tier Attributes” in the top toolbar. A window will appear in which you can specify the Participant, Annotator, and Content Language for a parent tier. Be sure to click “Change” before closing the window.

File-naming conventions

As with metadata, the best approach to file-naming conventions is to be proactive. One approach is to include the recording’s date at the start of the filename, in a year-month-day format so that the files can be sorted chronologically. A short informative tag, such as the narrative’s topic and/or a speaker’s initials, could follow the date.

- Example: 20250630-AJ.eaf could be the name for an ELAN annotation file corresponding to a recording with Alice Jones on June 30, 2025. Or 20250630-Alice-Jones.eaf could be used if Alice consents to having her name visible on the filename.

It is useful to use a shared filename for the ELAN file and its associated media file.

- Example: 202506030-AJ.eaf is the annotation file with transcriptions for the audio in 20250630-AJ.wav (or the video in 20250630.mp4)

Mobilizing transcriptions

Downloading our example files

This session of the workshop explores some ways to share already-created transcriptions for educational purposes. We will need a media file (file extension .mp4, .mp3, or .wav) and an associated ELAN annotation file (file extension .eaf).

1. Locate a media file and its associated (completed) ELAN annotation file. Save both files to the “ELAN Tutorial” folder in the “Desktop” directory.
2. Next, open the ELAN application. Instead of starting with a new transcription, go to **File > Open...** and select the .eaf file that you located in Step 1. The transcription will be loaded into the ELAN window.

For Resources 5 and 6, as well as the activity afterwards, we will be navigating in this ELAN window with the already-created transcription.

Resource 6: Exporting to text

To export to a text file (which you can then edit in other programs, such as Word, Googledocs, etc.), in the ELAN menu go to **File>>Export As>>Interlinear Text...**

1. This will bring up a window which will show your transcript file on the left, and various options for what you would like to appear in your exported file on the right
2. Make any updates to what you would like in your exported file, and then click **“Apply changes”** at the bottom of the window to preview what the exported file will look like (note updates to the text window on the left)
3. Once you are satisfied, click the **Save As** button at the bottom, and name and save your new text file -- now you can open it with another text editor, or save and send to someone
4. Some of the options you can select about *what* to show in the exported file:
 - Which tiers would you like to be visible in the exported file? (click boxes in the top right window)
 - Whether you want to see the tier “labels” (the names of the tiers), or not.
 - Whether you want to see the time-stamps and in what format (“show time code”)
5. Some of the options about *how* it should appear:
 - Width of text (in number of characters) before a line break
 - Spacing between the blocks of text

More information from the ELAN user guide:

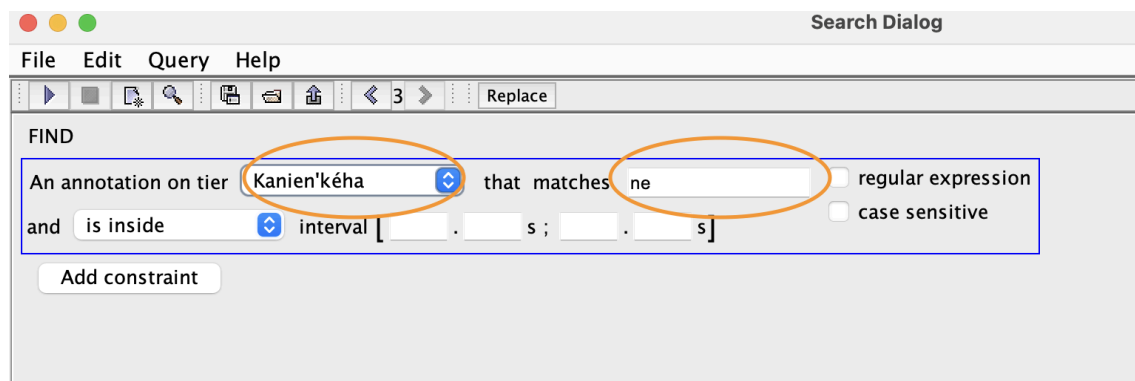
https://www.mpi.nl/corpus/html/elan/ch01s04.html#Sec_Exporting_to_interlinear_text

Resource 7: Corpus searches

The purpose of a corpus search is to retrieve *results* that match a *query*. This section is intended as an introduction, with many more features to explore.

Accessing the search function (for searching over a single ELAN file)

1. Open the ELAN file that you wish to search over.
2. In the upper toolbar, select “Search” and then “Find (and Replace)”. A new window appears in which you will enter and run your search query.
3. In the search window, add your keyword(s) to the field that appears to the right of “that matches”. For an example case where the ELAN file contains transcriptions of the language Kanien’kéha, we will look for instances of the Kanien’kéha particle “ne”.
4. Before running the search, we can (and often should) specify which ELAN tier to search. In our example, we will look at the tier called “Kanien’kéha” which contains Kanien’kéha orthography. In the drop-down menu to the right of “An annotation on tier”, select the appropriate tier name.



5. There are two ways to run the search: click the purple triangle in the upper-left corner, or click “Search” at the bottom-centre of the window.
6. After you run the search, you will see the results below the query, alongside associated information like the start- and stop-time for each segment that matches the query.

> Nr	Annotation
1	Ne tho ní:ioht tsi entkatáhsawen'.
2	Káti ne
3	ne shikeksà:'a, rake'nihkénha
4	tho nón: Oha'kwarónthne nónkwa ne Ahkwesáhsne
5	teiohnekónkies ne ohné:kanos ne
6	tánon óhionio - kahionió:kon, tho nítsi ki' tsi ensathón:te'ne'
7	ne ki kenh nón:
8	ne ohniare'kó:wa rón:kwe tehotténion."
9	"Weh" wa'ki:ron', "oh nonkié:re' ne tsi enhská:raton' tánon ó:nen ken' ítene's?"
10	Wahén:ron, ne's ki wahén:ron "iorihó:wanen ki' ki "
11	ne ki' ki
12	rateneri'tentóhrha tsi ne ki nahò:ten'k
13	nek ne rentóhrha
14	tánon ne rohtsi:'a kwáh ne tokén: tsi

55 occurrences in 45 annotations

- It is possible that the search results are too broad. For instance, one of the search results was *tánon óhionio - kahionió:kon, tho nítsi ki' tsi ensathón:te'ne'*... this annotation contains the characters “ne”. However, we are not looking for just any appearance of those two letters. Instead, we want to only find examples containing “ne” when it is represented as a standalone word.

Refining searches using regular expressions

- Regular expressions are a set of keyboard symbols that allow us to make more nuanced searches. The ELAN manual has more information at [https://www.mpi.nl/tools/elan/docs/manual/index.html#App Regular Expression.html](https://www.mpi.nl/tools/elan/docs/manual/index.html#App%20Regular%20Expression.html).
- We will use regular expressions to pinpoint “ne” as not just any string containing “ne” but specifically the standalone word.
- Returning to the example search, click on the text in the box below “FIND” at the top of the search window to resume editing the search query.
- Check the box for “regular expression” and add “\s” to both sides of the “ne” that you had previously written (quotation marks excluded).

Search Dialog

File Edit Query Help

Replace

FIND

An annotation on tier Kanien'kéha that matches \sne\s ☒ regular expression

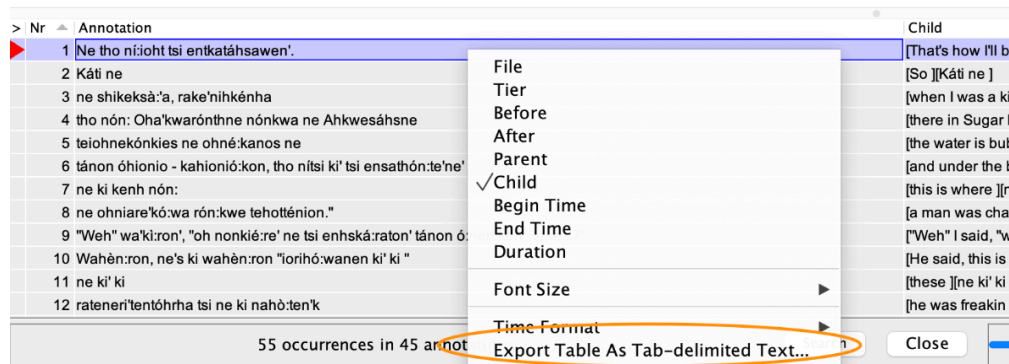
and is inside interval [. s ; . s] ☐ case sensitive

Add constraint

- Each “\s” indicates that a whitespace must be matched in the search results. In other words, running this search will only return “ne” that is surrounded on both sides by whitespace.
- Run the search again, and observe that all search results contain a standalone “ne”.

Exporting search results

- One way to save the output of search results is to create a tab-delimited text file. On Windows, right-click anywhere in the search results (except the column labels). On Mac, hold down the control key and click anywhere in the search results (except the column labels).
- Select the fields that you wish to be displayed in the text file by checking or un-checking the fields from the list.
- Select “Export Table As Tab De-limited Text..” and provide a name for the file. Select the default UTF-8 encoding.
- The new .txt file is most readable if you open it in a spreadsheet viewer, such as Microsoft Excel, Numbers, or Google Sheets.



Activity: Searching for keywords in a narrative

We will continue using the already-created transcription to learn about ELAN's search capabilities.

Independently or with a small group, come up with a list of keywords you'd like to search for in the transcription. Perhaps you can delegate different group members to different keywords

1. Using Resource 7 as a guide, search for your keyword as a plain keyword on the appropriate ELAN tier, without using regular expressions.
2. Again run a search for your particle on the appropriate ELAN tier, but now use regular expressions. A helpful regular expression is \W which stands for anything that is not an alphabetic character.
 - a. Example: if you run the search query \Wne\W then you will find all instances of “ne” that have no alphabetic characters on the left or on the right. Try out this strategy in replacing “ne” with your keyword!
3. Looking into your search results, discuss among your small group what you have found.
 - a. How many occurrences of the keyword did you find?
 - b. Are there particular words or phrases that your keyword tends to appear with?
 - c. Are there any instances of the keyword that surprised you?
 - d. Are you able to come up with an English translation of the keyword?
4. Export your search results to a .txt file.

Resource 8 (supplement): Creating your own ELAN template

Intro:

In order to create your own template, you will need to create a regular transcription file (*File > New*, add a media file, click ok). Once that's done, you will have to create types and tiers. These types and tiers are the important part in creating your template. Once you've created that template, you can simply save it as such and reuse it whenever it is needed. You'll find definitions for types below, and the set of steps needed to create types and tiers, and to save your file as a template.

Types:

ELAN has 4 basic types, which correspond to 4 types of relationships between tiers. These types specify for a given tier (i) whether it is time-aligned to the recording, (ii) whether it is related to another tier, and if it is related to another tier, in what way. Independent tiers have no type, and dependent tiers must have one of the 4 types. Here they are.

1. Time subdivision

- a. Annotations on this tier are time-aligned to the recording.
- b. Annotations on a *time subdivided* tier are always included in an annotation on the parent tier. They are smaller units of language included in the larger annotation on the parent tier.
- c. There are no gaps between those smaller units unless there is gap on the corresponding annotation on the parent tier
- d. Ideal to create annotations for single words that are time-aligned to those words in the recording.

2. Symbolic subdivision

- a. Annotations on this tier are not time-aligned to the recording.
- b. Annotations on a *symbolically-subdivided* tier are always included in an annotation on the parent tier. They are smaller units of language included in the larger annotation on the parent tier.
- c. There are no gaps between those smaller units unless there is a gap on the corresponding annotation on the parent tier.

- d. Ideal to create annotations for single morphemes (parts of words) that do not need to be time-aligned to those morphemes in the recording.
3. Included in
 - a. Annotations on this tier are time aligned to the recording.
 - b. Annotations on an *included-in* tier are always included in an annotation on the parent tier. They are smaller units of language included in the larger annotation on the parent tier.
 - c. There can be gaps between those smaller units of language.
 - d. Ideal to create annotations for words or sentences that are separated by pauses.
 4. Symbolic association
 - a. Annotations on this tier are not time-aligned to the recording.
 - b. Annotations on a *symbolically-subdivided* tier are always associated to an annotation on the parent tier of the same length
 - c. Ideal to create annotations for translations or comments.

Creating types

In the ELAN top menu, go to *Type > Add new type*

The 'Add Type' dialog box contains a table of 'Current Types' and a form to add a new type. The table lists various types and their properties. The form below the table has fields for 'Type Name', 'Stereotype', 'Use Controlled Vocabulary', 'Lexicon connection', 'ISO Data Category', and a 'Time-alignable' checkbox. Three red circles with numbers 1, 2, and 3 highlight the 'Type Name', 'Stereotype', and 'Time-alignable' fields respectively.

Type Name	Stereotype	Use Controlled ...	DC ID	Time-alignable
Syntax	Symbolic Association	-	-	yes
Motion Code	Symbolic Association	Motion	-	yes
M Expression	Time Subdivision	-	-	✓
P Expression	Time Subdivision	-	-	✓
Sp Comm	Symbolic Association	-	-	yes
Gesture #	-	-	-	✓

Form fields:

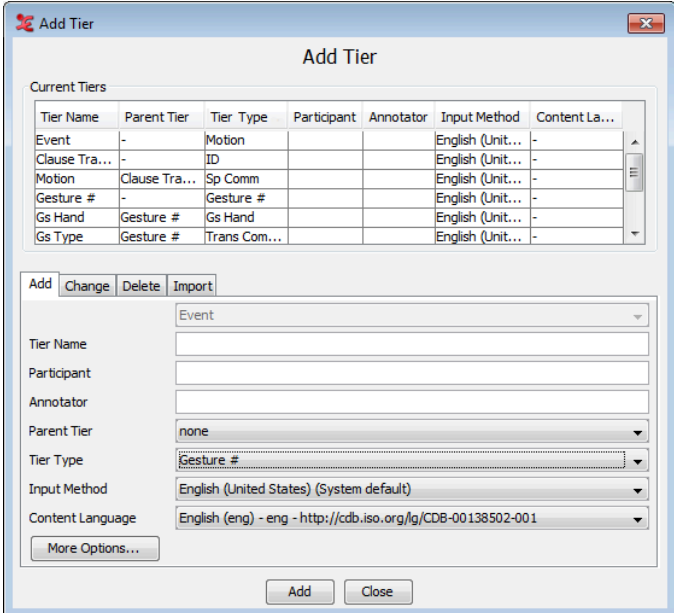
- Type Name: (highlighted with red circle 1)
- Stereotype: (highlighted with red circle 2)
- Use Controlled Vocabulary:
- Lexicon connection:
- ISO Data Category:
- Time-alignable: ☒ (highlighted with red circle 3)

1. Give your type a name that will be easy to remember. For instance, if you create a type for transcription tiers, you can simply call it *transcription*.
2. Choose the kind of type this tier should have, based on the description above. You can also follow the flow chart below.

3. Click *add*.
4. Repeat steps 1-3 for each type you need to create.
5. Note that you might need only 2 types even if you have 4 tiers for each speaker.

Creating tiers

- In the ELAN top menu, go to *Tier > Add new tier*
- Fill the dialog window below.



The 'Add Tier' dialog box contains a table of 'Current Tiers' and a form for adding a new tier.

Tier Name	Parent Tier	Tier Type	Participant	Annotator	Input Method	Content La...
Event	-	Motion			English (Unit...	-
Clause Tra...	-	ID			English (Unit...	-
Motion	Clause Tra...	Sp Comm			English (Unit...	-
Gesture #	-	Gesture #			English (Unit...	-
Gs Hand	Gesture #	Gs Hand			English (Unit...	-
Gs Type	Gesture #	Trans Com...			English (Unit...	-

Buttons: Add, Change, Delete, Import

Form fields:

- Event: Event
- Tier Name:
- Participant:
- Annotator:
- Parent Tier: none
- Tier Type: Gesture #
- Input Method: English (United States) (System default)
- Content Language: English (eng) - eng - <http://cdb.iso.org/l/g/CDB-00138502-001>

Buttons: More Options..., Add, Close

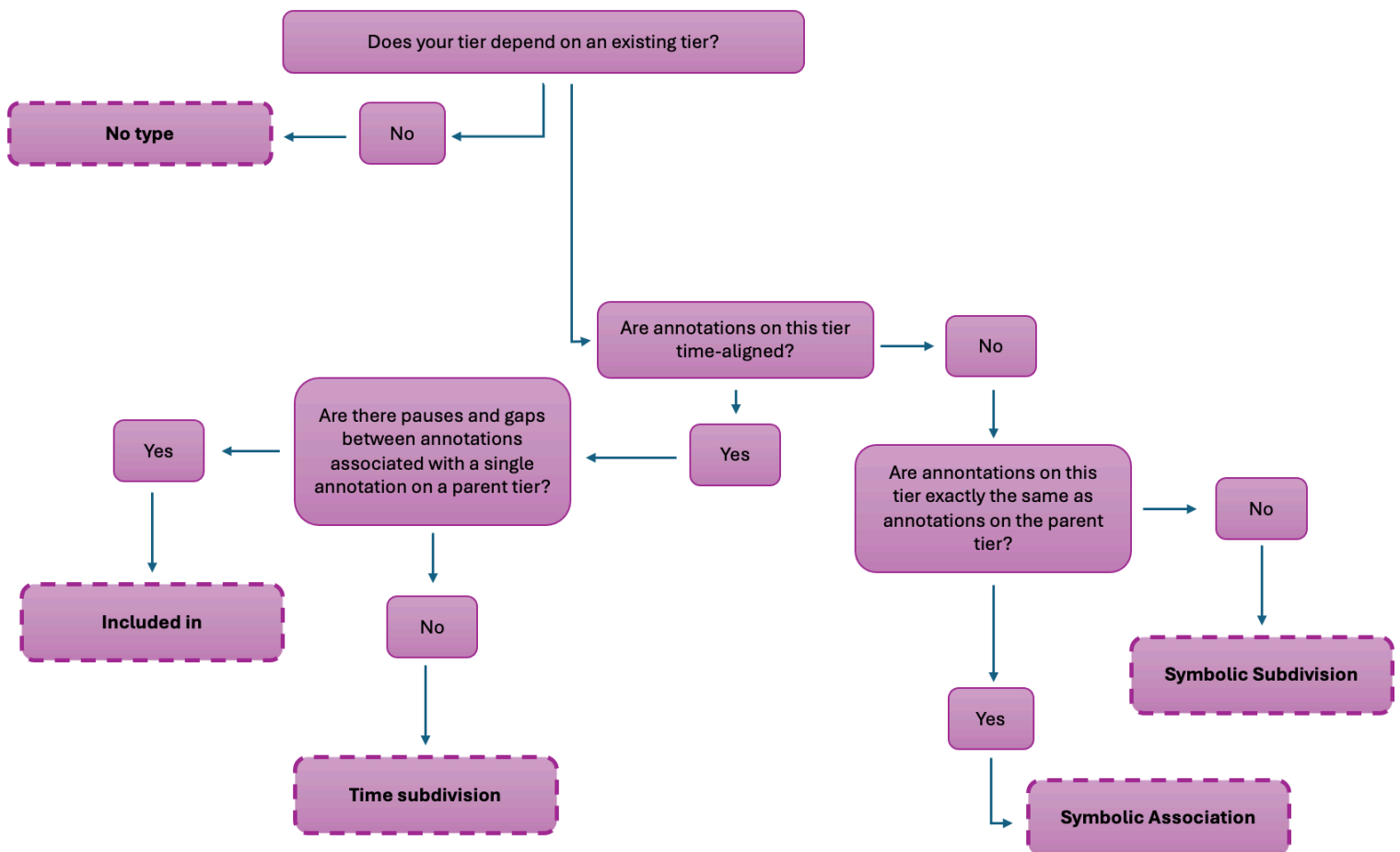
1. Input the **name** of your tier.
2. **Participant** refers to the speaker whose voice will be transcribed (optional).
3. **Annotator** refers to the person who is writing the annotations. (optional)
4. **Parent tier** refers to the tier on which this new tier is dependent. Usually, tiers where you transcribe the speaker's speech do not have parent tiers, but the tiers where you put transcriptions, notes, word-by-word parses and glosses are dependent tiers.
5. **Tier type** refers to the type that this new tier will have, it encodes the kind of relationship this tier has with a parent tier.
6. **Input method** refers to the set of characters you are using in your transcription. This is useful if you are using special characters that are associated with that language. (optional)
7. **Content Language** refers to the language you are transcribing. (optional)
8. Click *Add*.

9. Repeat steps 1-7 for each tier.

Creating your template

- Now that you have created the types and tiers you need, you can export the format of your file to be reused with as many files as you need. This will save you considerable time if you have to build a corpus (a set) containing multiple transcriptions.
- Go to *File > Save as template*
- This will generate an .etf file. This .etf file will only contain the tiers and types, no transcription, and you will be able to upload it whenever you are creating a new ELAN file.
- You can generate templates from any file you are using, so if you stumble upon a file that you find particularly suited to your needs, you can save it as a template for future uses.

How to choose your types - A flowchart

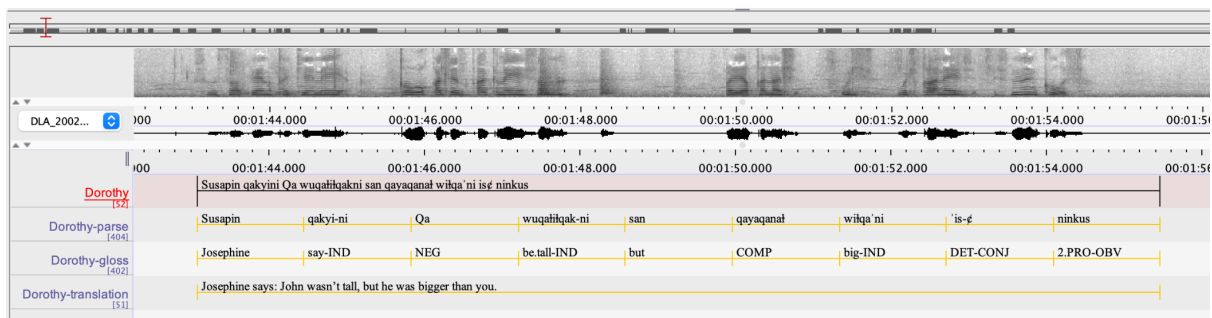


Resource 9 (supplement): Automatic morphemic tokenization

Intro:

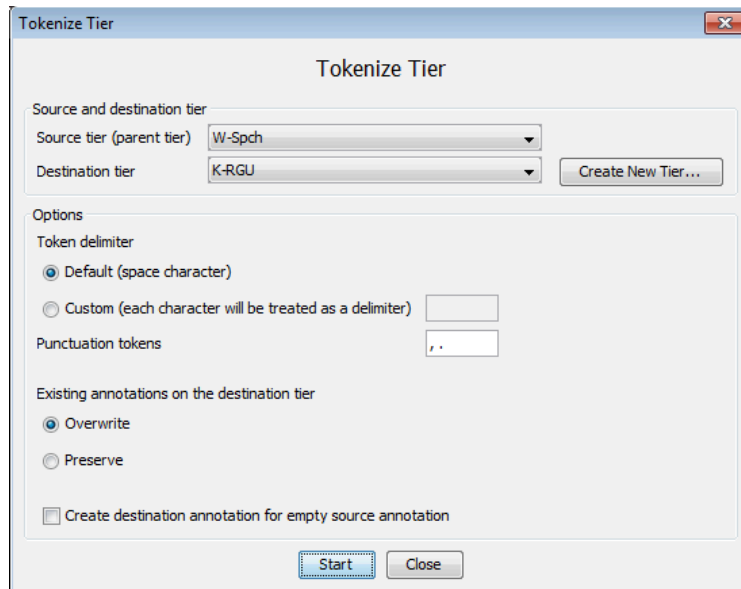
It is possible to ask ELAN to generate a set of annotations on a dependent tier automatically. This is useful if you need to generate an annotation for each word or each morpheme (part of word) in a tier. For instance, in the image below, the annotations on the tier called *Dorothy-parse* have been generated automatically by ELAN through a process called *tokenization*. When tokenizing a tier, ELAN looks for a specific character (called a delimiter) in a given tier. Whenever it encounters this character, it generates a new annotation on a dependent tier.

In the example below, ELAN tokenized the tier *Dorothy* to generate the annotations on the tier *Dorothy-parse*. Every time it encountered a space, it created a new annotation on the tier *Dorothy-parse*. As you can see in the example, each annotation on *Dorothy-parse* corresponds to a word.



- In the top menu, *Tier > Tokenize*
- In the dialog window, you'll have to specify
 - The source tier (the tier that will be tokenized)
 - The destination tier (you can also ask ELAN to automatically create a new tier containing the tokenized version of the parent tier).
 - a token delimiter, the specific character that ELAN should look for when tokenizing. By default, this is set to the space character, but you can choose something else.
 - If you already have annotations on the destination tier, you should click *Preserve*, and then click *start*.

(Continued on the next page)



- This process will generate a set of annotations that already contain text. This is useful if you want to then provide a word-for-word translation or glosses.