ELAN for APLS

Data Science for Linguists– Spring ‘24

Maya Asher
Working under Dr. Dan Villarreal

Table of contents

01 Background on APLS

02 Download and First Steps

03 Modes of Transcription

04 Shortcuts and Hotkeys

05 General Processes

01
Background on APLS

Background on APLS

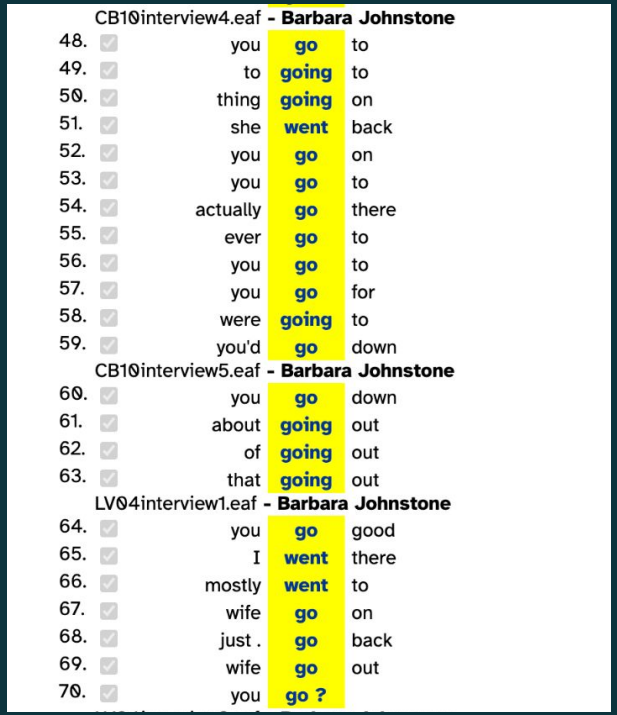
- Archive of Pittsburgh Language and Speech (APLS)
  - Database of time-aligned transcripts of audio files powered by LaBB-CAT
    - LaBB-CAT: open source software written by Robert Fromont and Jen Hay (NZILBB), made to store audio and visual recordings and their transcripts
  - Pittsburgh Speech and Society Project (PSSP)
    - Directed by Barbara Johnstone and Scott Kiesling
    - From 2003-2005
    - Sociolinguistic interviews with Pittsburgh residents from 4 different neighborhoods
  - Still under construction, but will become available to the public
    - ~24 hours uploaded out of ~48 hours

Potential Research with APLS

- Search features in APLS
  - View entire transcripts and download into many different formats (TextGrid, EAF, PDF, plain text, etc.)
  - Search for specific words, see them in context, and export as CSV/TextGrid
    - Can use regular expressions when searching (ex: .* for 0 or more of any character)
    - Can use many different projects/layers when searching (ex: lemma, morphology, phonology, syntax, etc.)
- Applications
  - Provides a large quantity of usable data for researchers looking for specific phonemes/morphemes/words used by Pittsburgh speakers

Potential Research with APLS

Lemma layer for word “go”



CB10interview4.eaf - **Barbara Johnstone**

48.	you	go	to
49.	to	going	to
50.	thing	going	on
51.	she	went	back
52.	you	go	on
53.	you	go	to
54.	actually	go	there
55.	ever	go	to
56.	you	go	to
57.	you	go	for
58.	were	going	to
59.	you'd	go	down

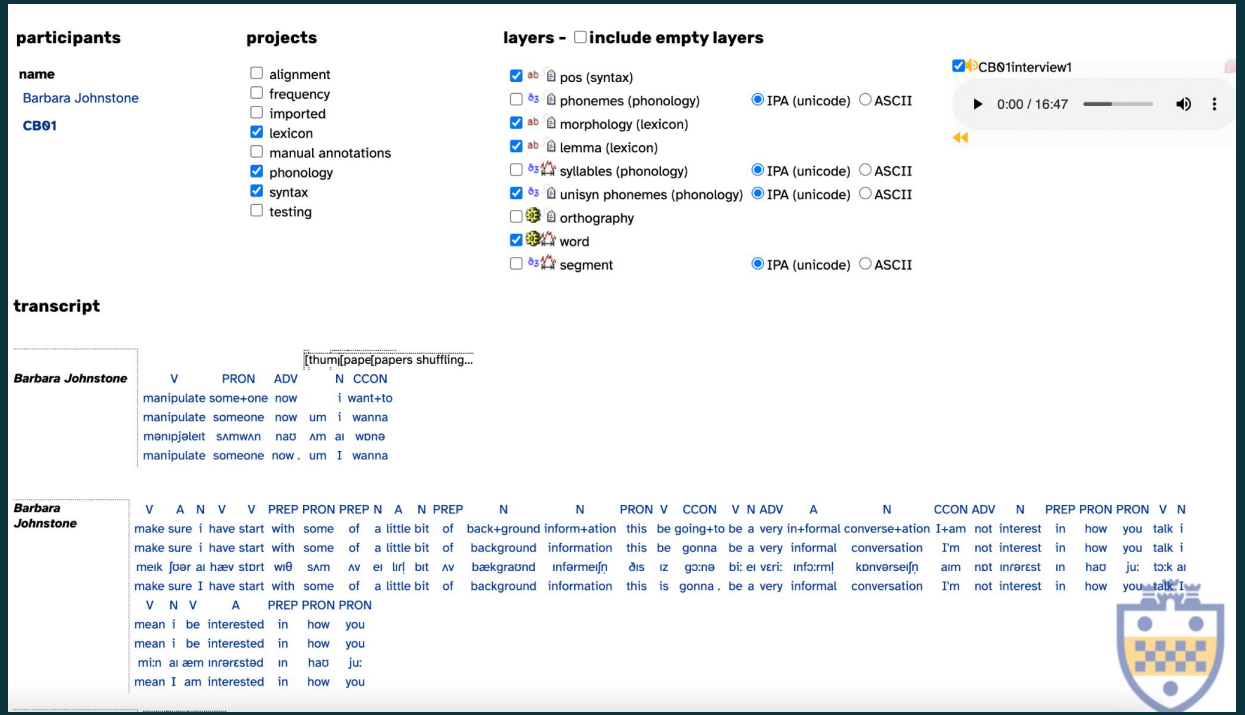
CB10interview5.eaf - **Barbara Johnstone**

60.	you	go	down
61.	about	going	out
62.	of	going	out
63.	that	going	out

LV04interview1.eaf - **Barbara Johnstone**

64.	you	go	good
65.	I	went	there
66.	mostly	went	to
67.	wife	go	on
68.	just .	go	back
69.	wife	go	out
70.	you	go ?	

Transcript with multiple layers



participants

name

Barbara Johnstone

CB01

projects

- alignment
- frequency
- imported
- lexicon
- manual annotations
- phonology
- syntax
- testing

layers - include empty layers

- pos (syntax)
- phonemes (phonology) — IPA (unicode) / ASCII
- morphology (lexicon)
- lemma (lexicon)
- syllables (phonology) — IPA (unicode) / ASCII
- unisyn phonemes (phonology) — IPA (unicode) / ASCII
- orthography
- word
- segment — IPA (unicode) / ASCII

CB01interview1 — 0:00 / 16:47

transcript

[thum][pape[papers shuffling...

Barbara Johnstone

V	PRON	ADV	N	CCON
manipulate	some+one	now	i	want+to
manipulate	someone	now	um	i wanna
mənɪpjəleɪt	sʌmwʌn	naʊ	ʌm	aɪ wɒnə
manipulate	someone	now .	um	I wanna

Barbara Johnstone

V	A	N	V	V	PREP	PRON	PREP	N	A	N	PREP	N	N	PRON	V	CCON	V	N	ADV	A	N	CCON	ADV	N	PREP	PRON	PRON	V	N
make	sure	i	have	start	with	some	of	a	little	bit	of	back+ground	inform+ation	this	be	going+to	be	a	very	in+formal	converse+ation	I+am	not	interest	in	how	you	talk	i
make	sure	i	have	start	with	some	of	a	little	bit	of	background	information	this	be	gonna	be	a	very	informal	conversation	I'm	not	interest	in	how	you	talk	i
meɪk	ʃʊər	aɪ	hæv	stɔrt	wɪθ	sʌm	ʌv	eɪ	lɪɾl	bɪt	ʌv	bækgraʊnd	ɪnfɚmeɪʃn̩	ðɪs	ɪz	gɔːnə	biː	eɪ	vɛriː	ɪnfɔːrml̩	kɒnvɚseɪʃn̩	aɪm	nɒt	ɪnɹɚɹɛst	ɪn	haʊ	juː	tɔːk	aɪ
make	sure	I	have	start	with	some	of	a	little	bit	of	background	information	this	is	gonna .	be	a	very	informal	conversation	I'm	not	interest	in	how	you	talk	I

V	N	V	A	PREP	PRON	PRON
mean	i	be	interested	in	how	you
mean	i	be	interested	in	how	you
miːn	aɪ	æm	ɪnɹɚɹɛstɪd	ɪn	haʊ	juː
mean	I	am	interested	in	how	you

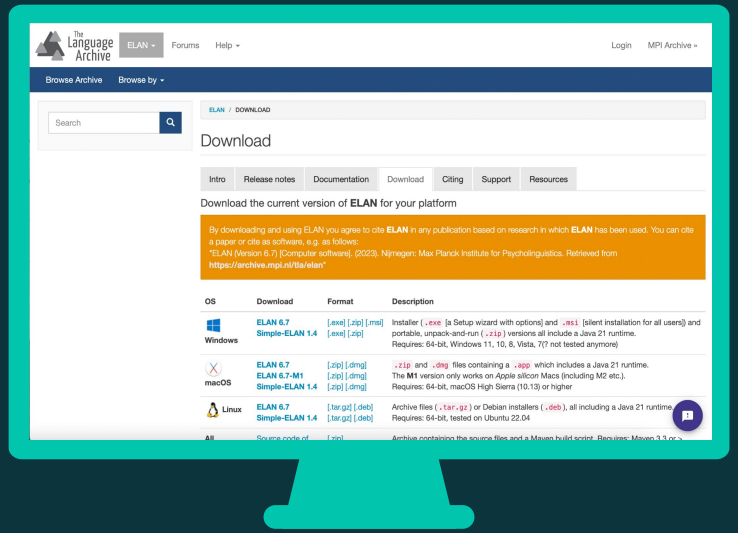
Recent Advances in Our Transcription

- The Beautiful Monster
  - Combines ASR and STT to output a fully segmented and transcribed file!
    - ASR: an AI segmented file (courtesy of Alejandro Ciuba)
    - STT: Batchalign file (courtesy of Houjun Liu, a freshman at Stanford :0)
- ELAN File Checker ✔️
  - Input a completed ELAN file and it checks for…
    - Valid tier names and attributes
    - Out-of-dictionary words
    - Overlaps
  - If all of those come up clean, it outputs an ELAN file with aligned segment boundaries

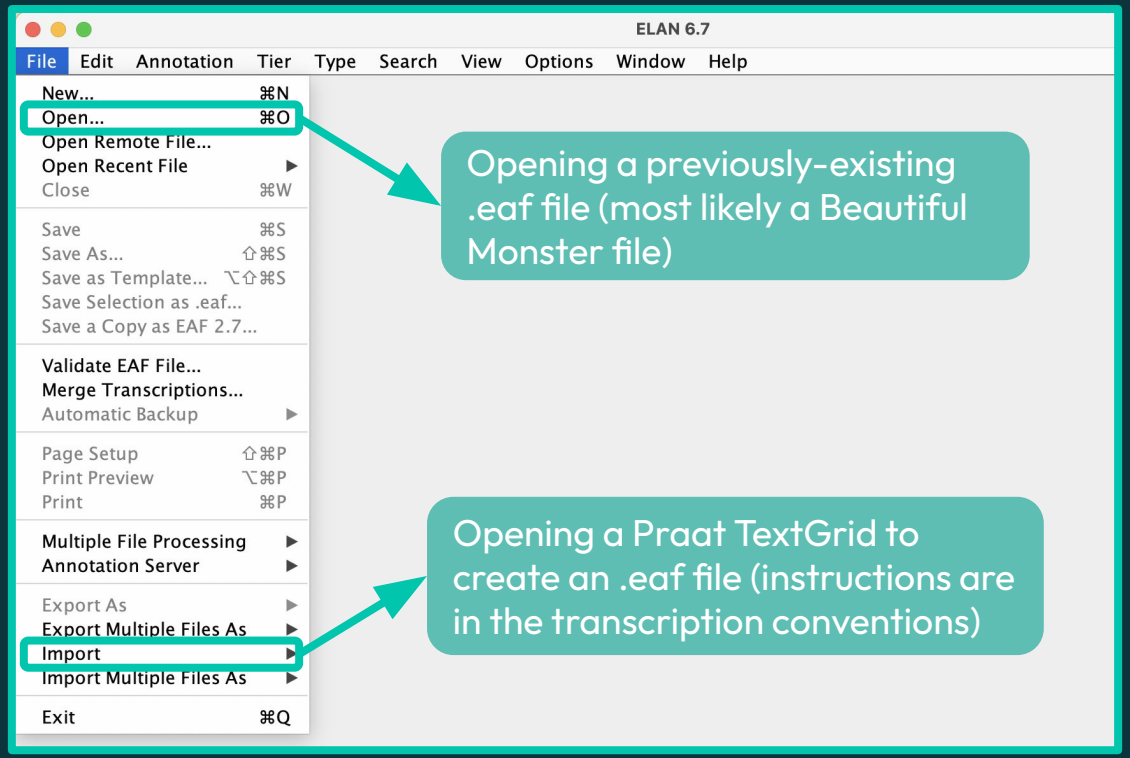
02
Download and First Steps

Download

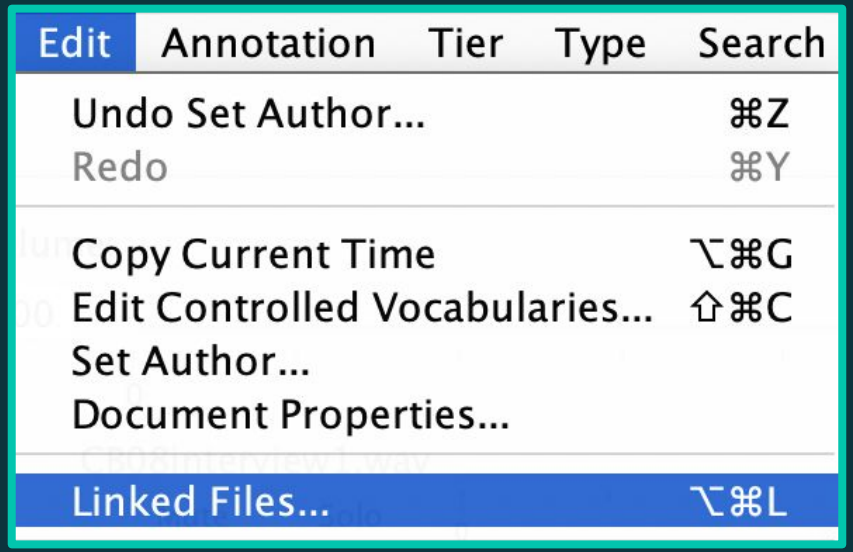
Get ELAN for free on the internet:
https://archive.mpi.nl/tla/elan/download



First Steps

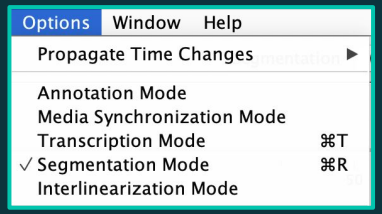

ELAN 6.7
File
Edit
Annotation
Tier
Type
Search
View
Options
Window
Help
New... ⌘N
Open... ⌘O
Open Remote File...
Open Recent File
Close ⌘W
Save ⌘S
Save As... ⇧⌘S
Save as Template... ⌥⇧⌘S
Save Selection as .eaf...
Save a Copy as EAF 2.7...
Validate EAF File...
Merge Transcriptions...
Automatic Backup
Page Setup ⇧⌘P
Print Preview ⌥⌘P
Print ⌘P
Multiple File Processing
Annotation Server
Export As
Export Multiple Files As
Import
Import Multiple Files As
Exit ⌘Q
Opening a previously-existing .eaf file (most likely a Beautiful Monster file)
Opening a Praat TextGrid to create an .eaf file (instructions are in the transcription conventions)

Add Audio


Edit
Annotation
Tier
Type
Search
Undo Set Author... ⌘Z
Redo ⌘Y
Copy Current Time ⌥⌘G
Edit Controlled Vocabularies... ⇧⌘C
Set Author...
Document Properties...
Linked Files... ⌥⌘L

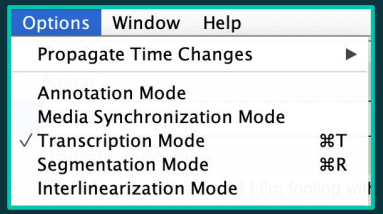
03
Modes of Transcription

3 Modes of Transcription



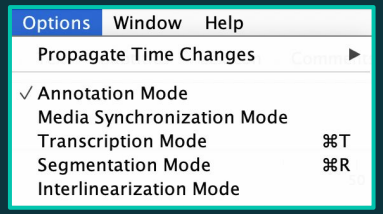
Segmentation

Used for creating and adjusting segments on all tiers



Transcription

Used for inputting transcriptions into segments



Annotation

Used for editing transcriptions directly on segments

We only really use these two

Segmentation Mode

Slider can slow down and speed up rate of speech

Play and pause button, among other things

The two channels: top is usually interviewer, bottom is usually interviewee

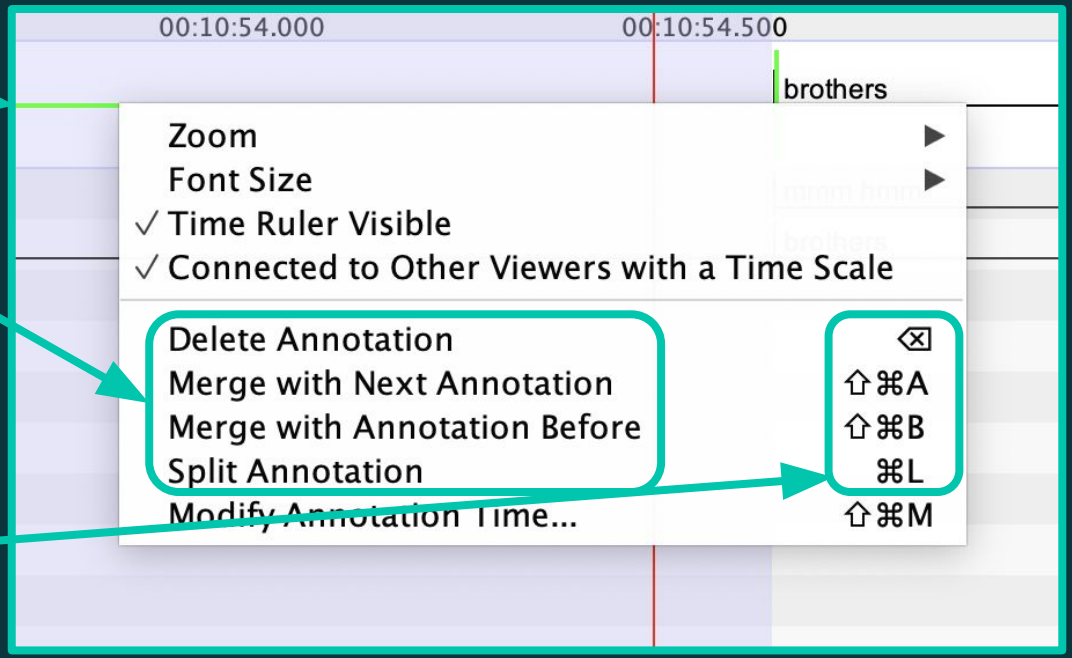
You can change current tier by double clicking the one you want to work with (or by using the up and down arrows)

Segmentation Mode

Right-click on the segment (it HAS to turn green)

These are the 4 you will use most often

Can also use hotkeys…but we have had issues with them in the past



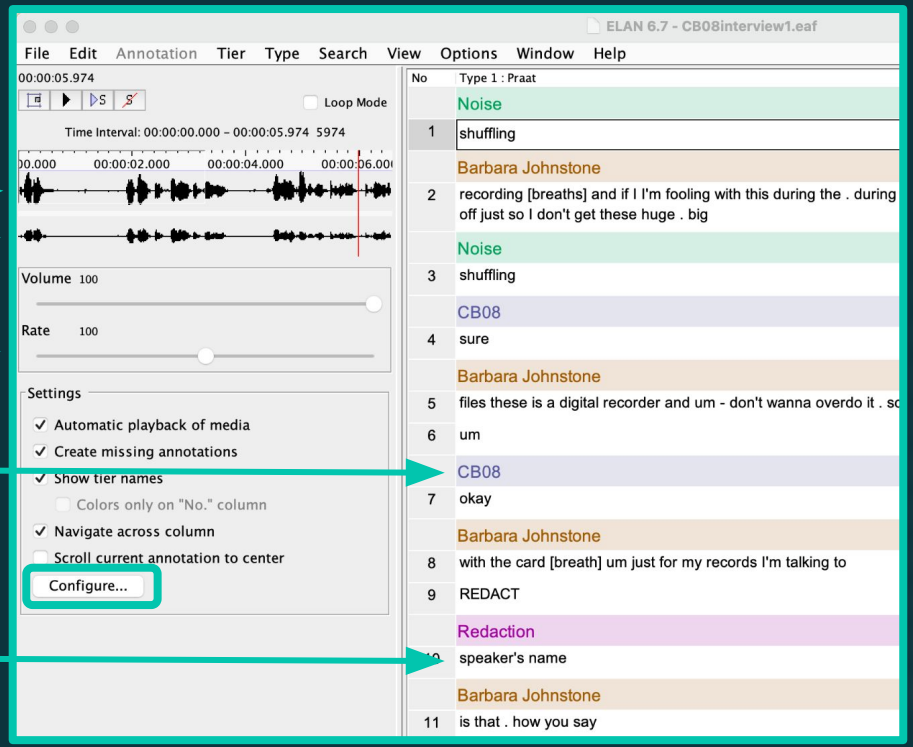
Transcription Mode

The two channels

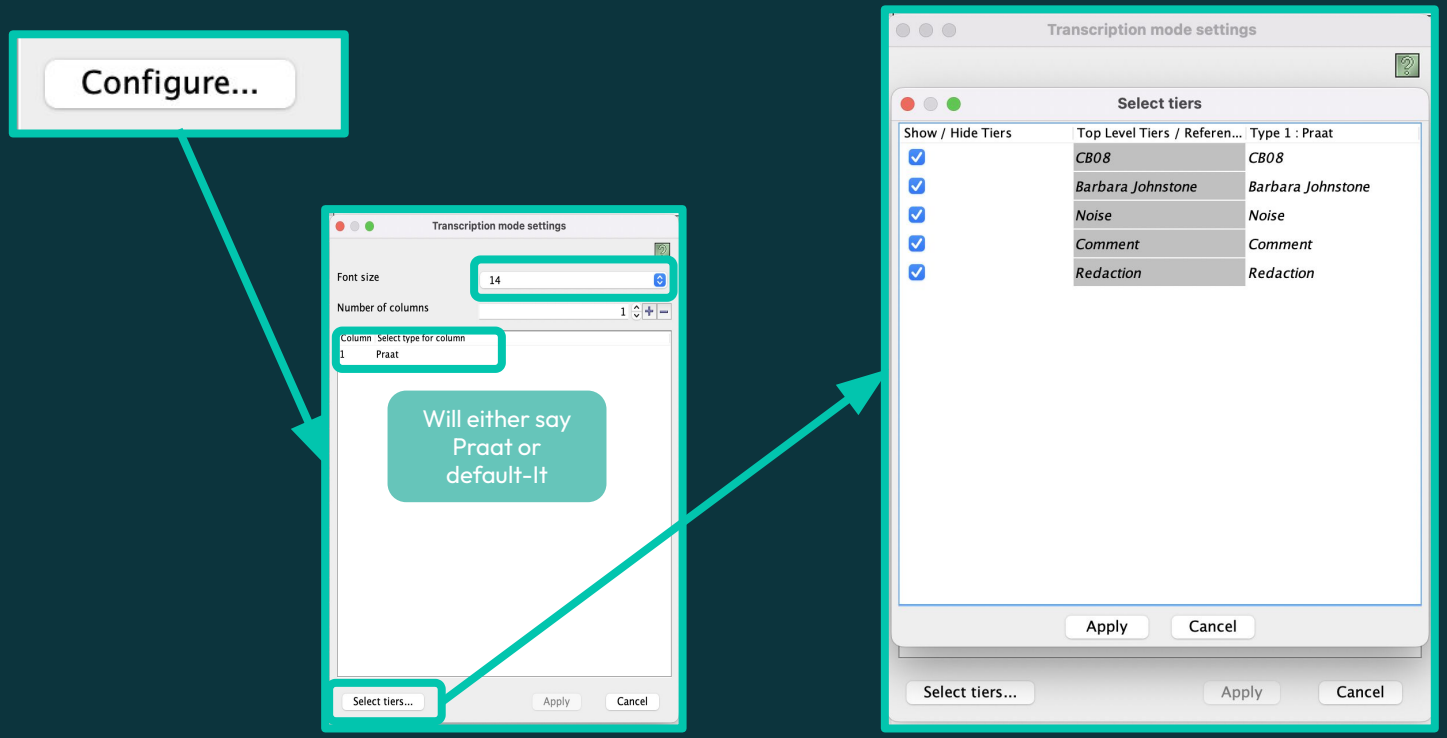
Slider can slow down and speed up rate of speech

Tier name

Transcription inside segment



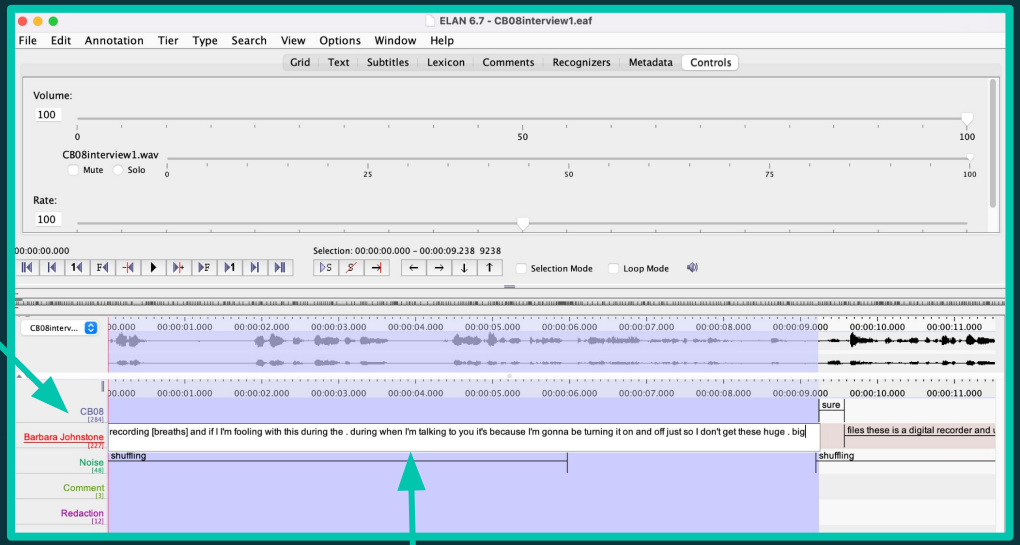
Transcription Mode



Annotation Mode

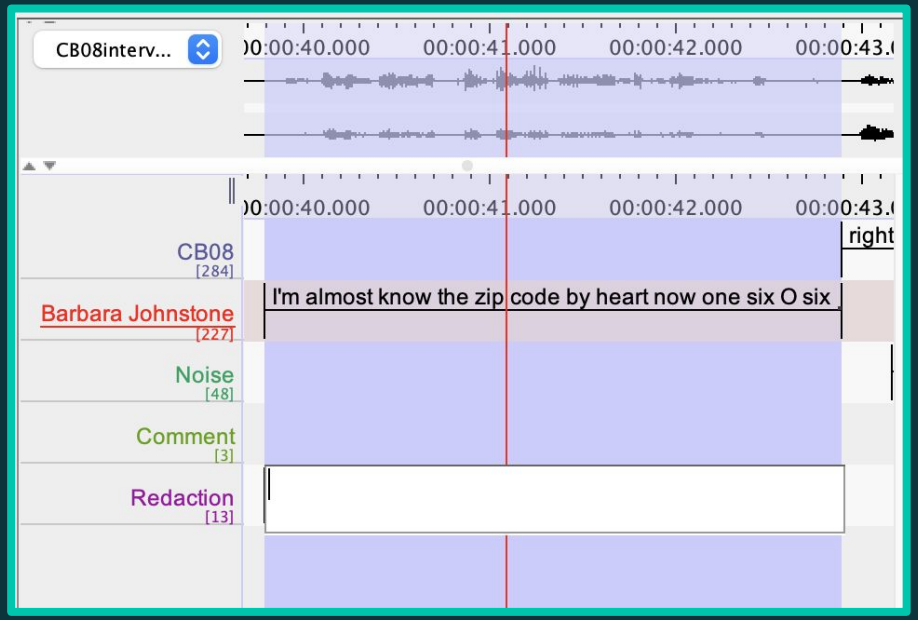
Shows number of segments on each tier

Double click the tier you want to work on



Can’t adjust segments but can edit text inside the segments by double clicking them

Annotation Mode

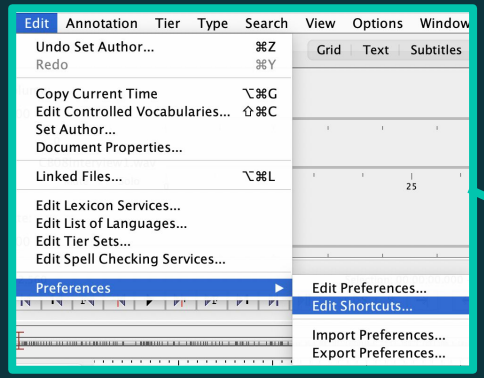


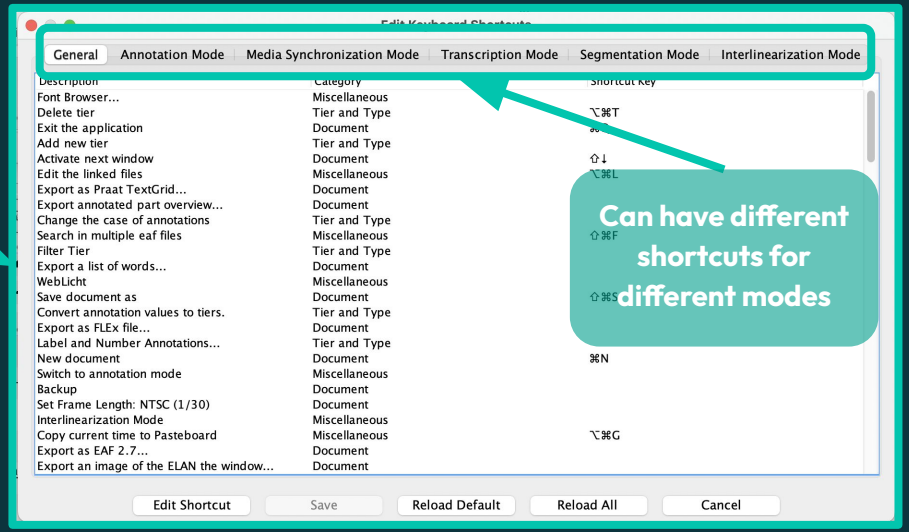
Copy segments onto other tiers:

1. Choose the tier that currently has the segment
2. Click segment until it is highlighted in purple
3. Double click on the tier that you want the new segment on
4. *Optional* write in the transcription directly

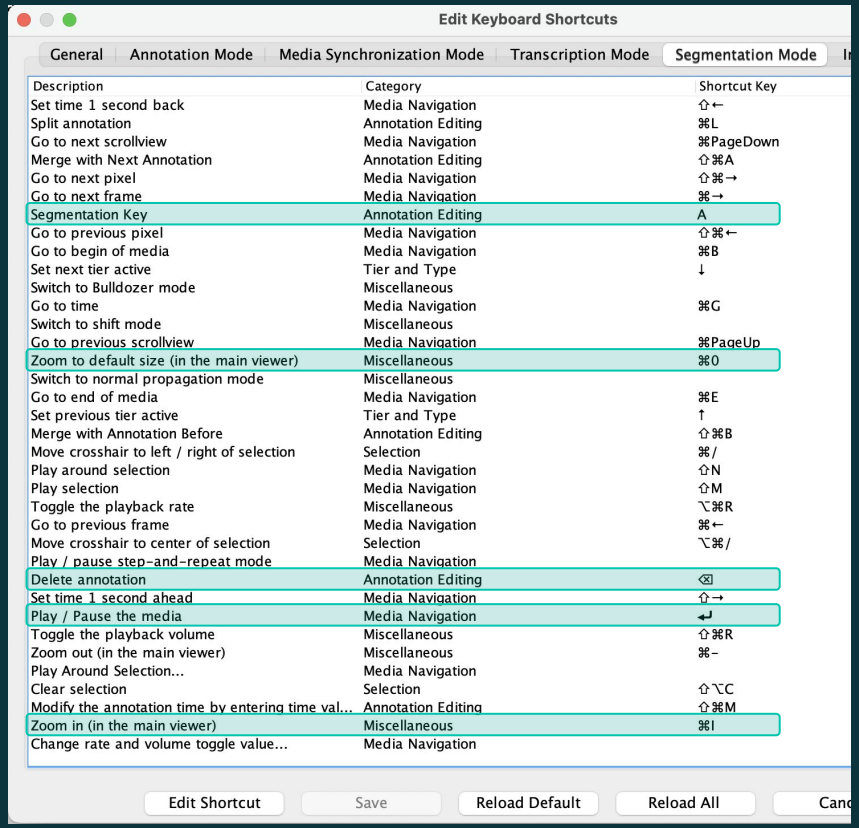
04
Shortcuts and Hotkeys

Shortcuts


Edit
Annotation
Tier
Type
Search
View
Options
Window
Undo Set Author... ⌘Z
Redo ⌘Y
Copy Current Time ⌥⌘G
Edit Controlled Vocabularies... ⇧⌘C
Set Author...
Document Properties...
Linked Files... ⌥⌘L
Edit Lexicon Services...
Edit List of Languages...
Edit Tier Sets...
Edit Spell Checking Services...
Preferences
Edit Preferences...
Edit Shortcuts...
Import Preferences...
Export Preferences...
Grid
Text
Subtitles
25

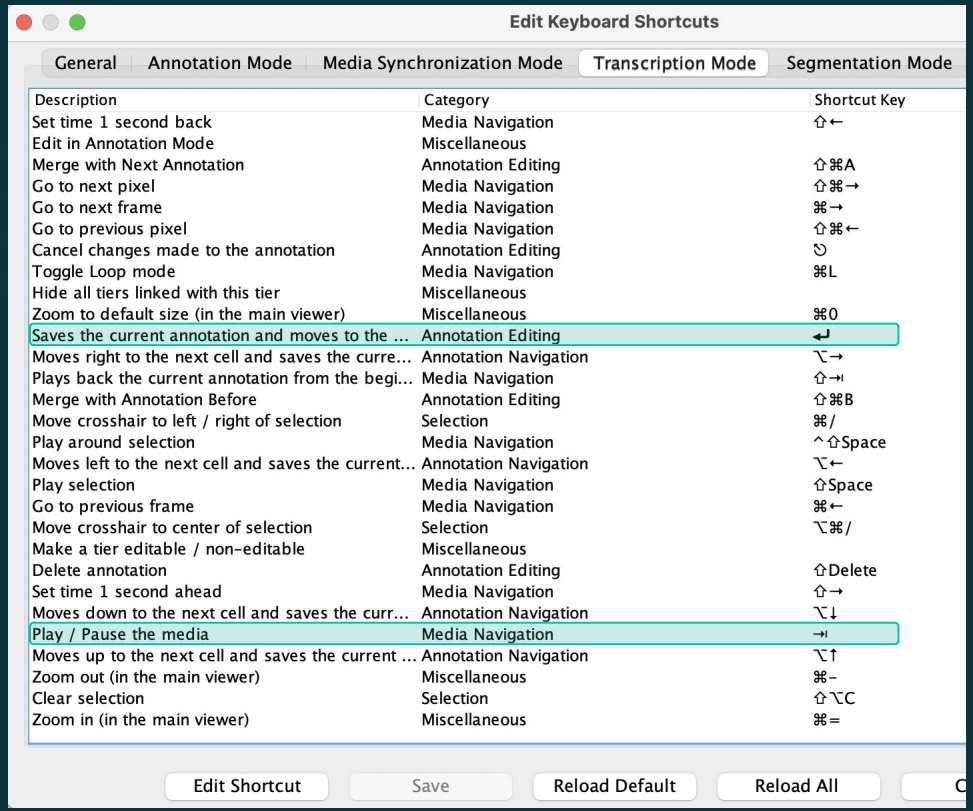

General
Annotation Mode
Media Synchronization Mode
Transcription Mode
Segmentation Mode
Interlinearization Mode
Description
Category
Shortcut Key
Font Browser... Miscellaneous
Delete tier Tier and Type ⌥⌘T
Exit the application Document
Add new tier Tier and Type
Activate next window Document ⇧↓
Edit the linked files Miscellaneous ⌥⌘L
Export as Praat TextGrid... Document
Export annotated part overview... Document
Change the case of annotations Tier and Type
Search in multiple eaf files Miscellaneous ⇧⌘F
Filter Tier Tier and Type
Export a list of words... Document
WebLicht Miscellaneous
Save document as Document ⇧⌘S
Convert annotation values to tiers. Tier and Type
Export as FLEx file... Document
Label and Number Annotations... Tier and Type
New document Document ⌘N
Switch to annotation mode Miscellaneous
Backup Document
Set Frame Length: NTSC (1/30) Document
Interlinearization Mode Miscellaneous
Copy current time to Pasteboard Miscellaneous ⌥⌘G
Export as EAF 2.7... Document
Export an image of the ELAN the window... Document
Edit Shortcut
Save
Reload Default
Reload All
Cancel
Can have different shortcuts for different modes

My Shortcuts: Segmentation



Segmentation Key	A
Delete Annotation	Delete
Zoom Out	Command Zero
Zoom In	Command I
Play/Pause	Return

My Shortcuts: Transcription



Description	Category	Shortcut Key
Set time 1 second back	Media Navigation	⇧←
Edit in Annotation Mode	Miscellaneous	
Merge with Next Annotation	Annotation Editing	⇧⌘A
Go to next pixel	Media Navigation	⇧⌘→
Go to next frame	Media Navigation	⌘→
Go to previous pixel	Media Navigation	⇧⌘←
Cancel changes made to the annotation	Annotation Editing	⎋
Toggle Loop mode	Media Navigation	⌘L
Hide all tiers linked with this tier	Miscellaneous	
Zoom to default size (in the main viewer)	Miscellaneous	⌘0
Saves the current annotation and moves to the ...	Annotation Editing	↵
Moves right to the next cell and saves the curre...	Annotation Navigation	⌥→
Plays back the current annotation from the begi...	Media Navigation	⇧⇥
Merge with Annotation Before	Annotation Editing	⇧⌘B
Move crosshair to left / right of selection	Selection	⌘/
Play around selection	Media Navigation	^⇧Space
Moves left to the next cell and saves the current...	Annotation Navigation	⌥←
Play selection	Media Navigation	⇧Space
Go to previous frame	Media Navigation	⌘←
Move crosshair to center of selection	Selection	⌥⌘/
Make a tier editable / non-editable	Miscellaneous	
Delete annotation	Annotation Editing	⇧Delete
Set time 1 second ahead	Media Navigation	⇧→
Moves down to the next cell and saves the curr...	Annotation Navigation	⌥↓
Play / Pause the media	Media Navigation	⇥
Moves up to the next cell and saves the current ...	Annotation Navigation	⌥↑
Zoom out (in the main viewer)	Miscellaneous	⌘-
Clear selection	Selection	⇧⌥C
Zoom in (in the main viewer)	Miscellaneous	⌘=

Move to next segment	Return
Play/Pause	Tab

THE MOST IMPORTANT SHORTCUT:

Control/Command S

05
General Processes

My Process

I fix the segments using *Segmentation Mode*

Segmentation

I usually repeat these two steps in 2-5 minute increments

Transcription

I input the transcription using *Transcription Mode*

I listen through the entire file and segment on the Noise tier using *Segmentation Mode*

Noises

Rechecks

I use the ELAN File Checker to clean up the file and find out-of-dictionary words

Thanks!

Questions?

Sources and Helpful Links

LaBB-CAT. (n.d.). SourceForge. Retrieved November 29, 2022, from https://onzeminer.sourceforge.net/#:~:text=LaBB%2DCAT%20is%20a%20browser,automatically%20generated%20or%20manually%20added

New Zealand Institute of Language, Brain and Behavior. (2003-2022). *LaBB-CAT* (20221013.1114). University of Canterbury, New Zealand. https://labb-cat.linguistics.pitt.edu/labbcat/

University of Pittsburgh. (n.d.). *Pittsburgh Speech & Society*. Research. https://pittsburghspeech.pitt.edu/index.html

ELAN:

https://archive.mpi.nl/tla/elan/download

Batchalign:

https://github.com/talkbank/batchalign2

PSSP:

https://pittsburghspeech.pitt.edu/

Alejandro Ciuba's Website:

https://alejandrociuba.github.io/

Houjun Liu's LinkedIn:

https://www.linkedin.com/in/houjun-liu/