



ELAN for Transcription

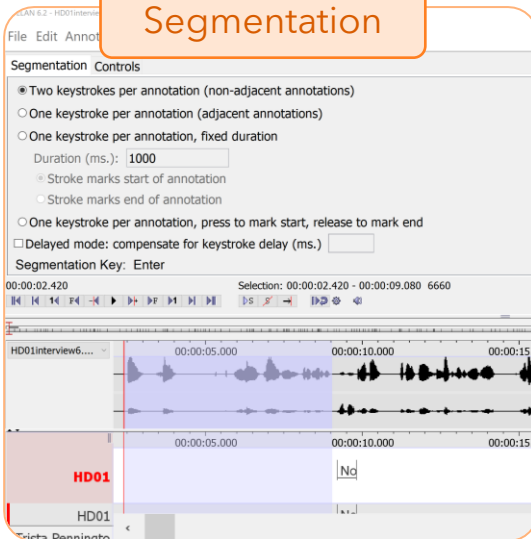
Emma McKibbin



What is ELAN?

- **EUDICO Linguistic Annotator**
- Max Planck Institute for Psycholinguistics
 - Brugman (2004), Wittenburg et al. (2006)
- First released in 2000 as EAT (EUDICO Annotation Tool)
 - Renamed ELAN in 2002
- Free, open source

Segmentation



ELAN 6.2 - HD01interview6...

File Edit Annot...

Segmentation Controls

- Two keystrokes per annotation (non-adjacent annotations)
- One keystroke per annotation (adjacent annotations)
- One keystroke per annotation, fixed duration

Duration (ms.): 1000

- Stroke marks start of annotation
- Stroke marks end of annotation

- One keystroke per annotation, press to mark start, release to mark end

Delayed mode: compensate for keystroke delay (ms.)

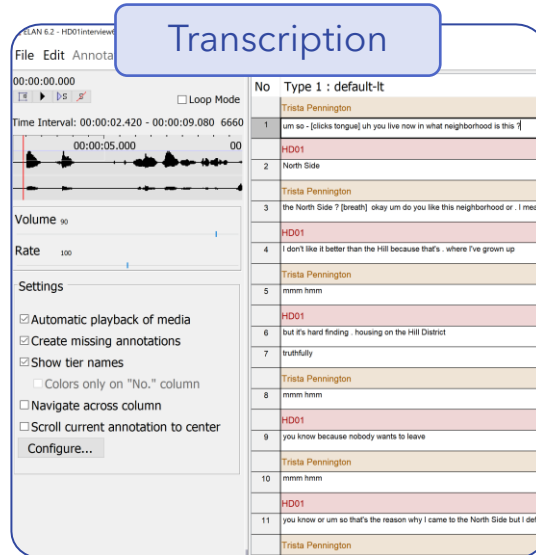
Segmentation Key: Enter

00:00:02.420 Selection: 00:00:02.420 - 00:00:09.080 6660

HD01

Trista Pennington

Transcription



ELAN 6.2 - HD01interview6...

File Edit Annota...

00:00:00.000

Time Interval: 00:00:02.420 - 00:00:09.080 6660

00:00:05.000

Volume 90

Rate 100

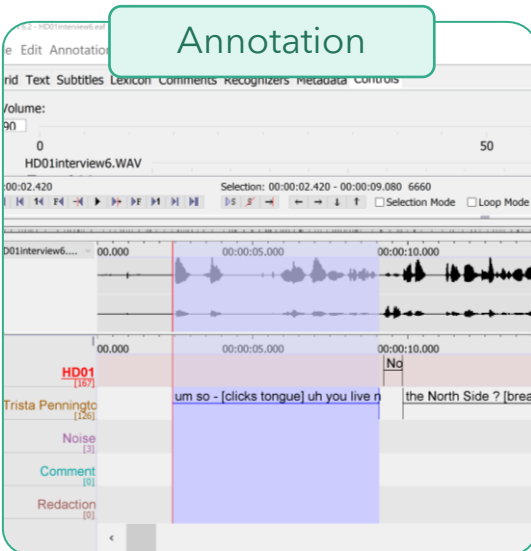
Settings

- Automatic playback of media
- Create missing annotations
- Show tier names
 - Colors only on "No." column
- Navigate across column
- Scroll current annotation to center

Configure...

No	Type 1 : default-It
1	Trista Pennington
	um so - [clicks tongue] uh you live now in what neighborhood is this {
HD01	
2	North Side
Trista Pennington	
3	the North Side ? [breath] okay um do you like this neighborhood or, I mean
HD01	
4	I don't like it better than the Hill because that's, where I've grown up
Trista Pennington	
5	mum hnsn
HD01	
6	but it's hard finding, housing on the Hill District
Trista Pennington	
7	truthfully
Trista Pennington	
8	mum hnsn
HD01	
9	you know because nobody wants to leave
Trista Pennington	
10	mum hnsn
HD01	
11	you know or um so that's the reason why I came to the North Side but I defin
Trista Pennington	

Annotation



ELAN 6.2 - HD01interview6...

File Edit Annota...

Volume: 0 50

HD01interview6.WAV

00:00:02.420 Selection: 00:00:02.420 - 00:00:09.080 6660

HD01

Trista Pennington (187)

Noise (3)

Comment (0)

Redaction (0)

um so - [clicks tongue] uh you live n the North Side ? [breath

What is ELAN for?

- Audio and video annotation
 - Praat - audio ONLY!
- More flexible annotation interface than Praat
 - Segmentation Mode
 - Annotation Mode
 - Transcription Mode
- Not focused on media manipulation

Creating and Opening Files



Starting a new file:

- (1) Select a media file (.wav, .mp4, .mov, etc.)
- (2) Once media is open, you can add annotations and save.



Opening existing file in ELAN:

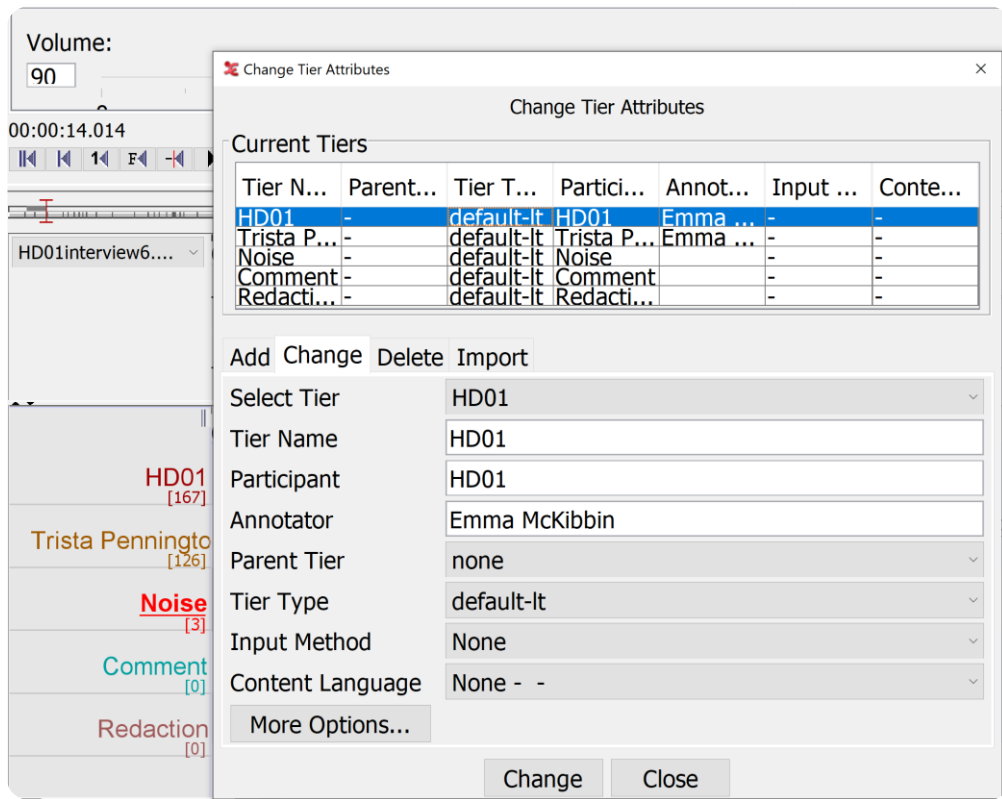
- (1) Just open the .eaf! Audio opens automatically

Annotations saved in .eaf format (XML)

- Media file unchanged
- .eaf linked to a media file via absolute path

```
MEDIA_URL="file:///C:/Users/Emma/Documents/Research/HD06-1.wav"
```

Recommend that you turn on **auto-saving**, since ELAN crashes sometimes!



Tiers

- Annotation layers
 - Similar to Praat tiers
- Hierarchical
 - Parent tier and referring tiers
 - E.g., you can annotate annotations
- Can include metadata
 - Annotator name
 - Participant name
 - Content language, etc.
- Color customization!
- Active tier is **red**

Using ELAN for Transcription Projects

General procedure:

1. Segment the file
2. Transcribe the file
3. Second-pass the file to correct errors
4. Ready for corpus / analysis!

But before any of this...

What guidelines will you follow to keep transcription consistent?

Transcription Conventions

Set by head researcher / project leader

Depend on project's end goals!

Things to consider...

- Non-prescriptive/non-standard forms
- Sensitive information
- Background noise and non-linguistic speaker noise
- Speech errors, incomplete or interrupted tokens
- Precision of segmentation and pauses
- Filler words
- Speaker overlap, if conversational
- Phonetic/phonemic vs. Orthographic transcription
- Punctuation (, . ' ? ! ")

	Barbara Johnstone
148	mmm hmm
	CB01
149	that's not his . primary . job
	Barbara Johnstone
150	[breath] now what about your kids ?
	CB01
151	[clicks tongue]
	Barbara Johnstone
152	what
153	do they do ?
	CB01
154	my son's a teacher . um . and my daughter's a speech therapist
	Barbara Johnstone
	mmm you said and and she's in Atlanta ?

1 File setup

1.1 Tips

2 Segmentation

2.1 Turns

2.2 Overlaps

2.3 Redaction

3 Annotation

3.1 Spelling

3.2 Punctuation

3.3 Comments and noises

3.4 Unfinished words and novel words



Transcription Workflow

STEP 1: Segmentation

Segmentation Mode:

- One tier at a time
- Keystroke controlled
- *Not great for precision - relies on reaction time!*

VS.

Annotation Mode:

- Select area and tier to create or adjust segment
- Transcribe in the segment, right below the sound wave
- *More control over segment boundaries, annotations editable!*

STEP 2: Transcription

Annotation Mode:

- Type directly on segments
- Overlapping segments visible
- *Longer segments difficult to transcribe/review due to window size*

VS.

Transcription Mode:

- View all turns chronologically (transcript only)
- Easier to see typos / if turns are attributed to correct speaker
- *Overlapping segments harder to identify visually*

Essentially... Annotation Mode is Praat.

But:

- More idiosyncratic keyboard shortcuts (by default)
- No spectrogram
- Space *between* segments does **not** count as a turn

Suggested Workflow:

- (1) Segment and first-pass transcribe in **Annotation Mode**
- (2) Check for errors in **Transcription Mode**



Thank You!

? - Feel free to ask questions - ?