Introduction to Elan

Nick Thieberger
Research Unit for Indigenous Language
School of Languages and Linguistics
The University of Melbourne

ARC Centre of Excellence for the Dynamics of Language

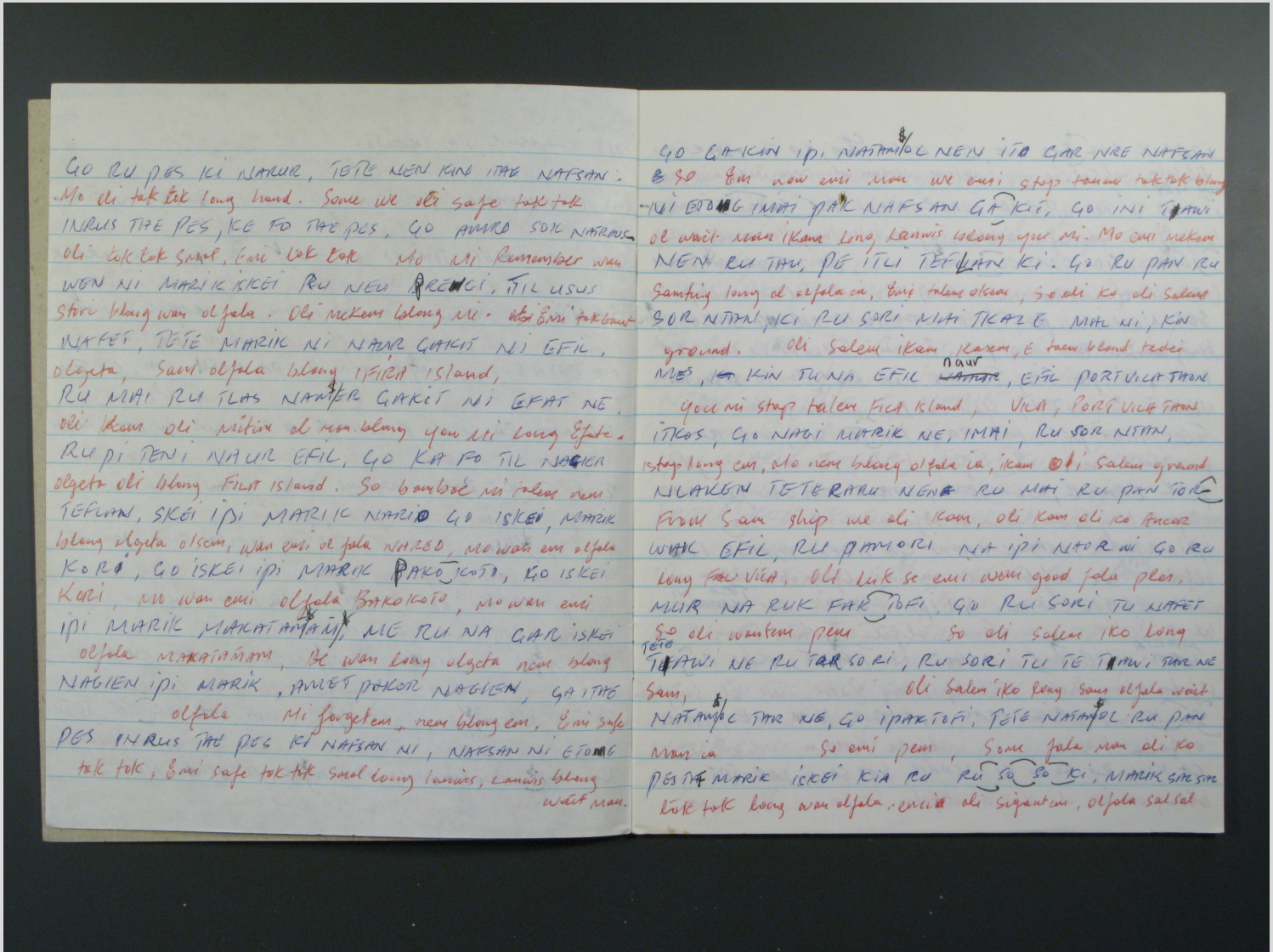


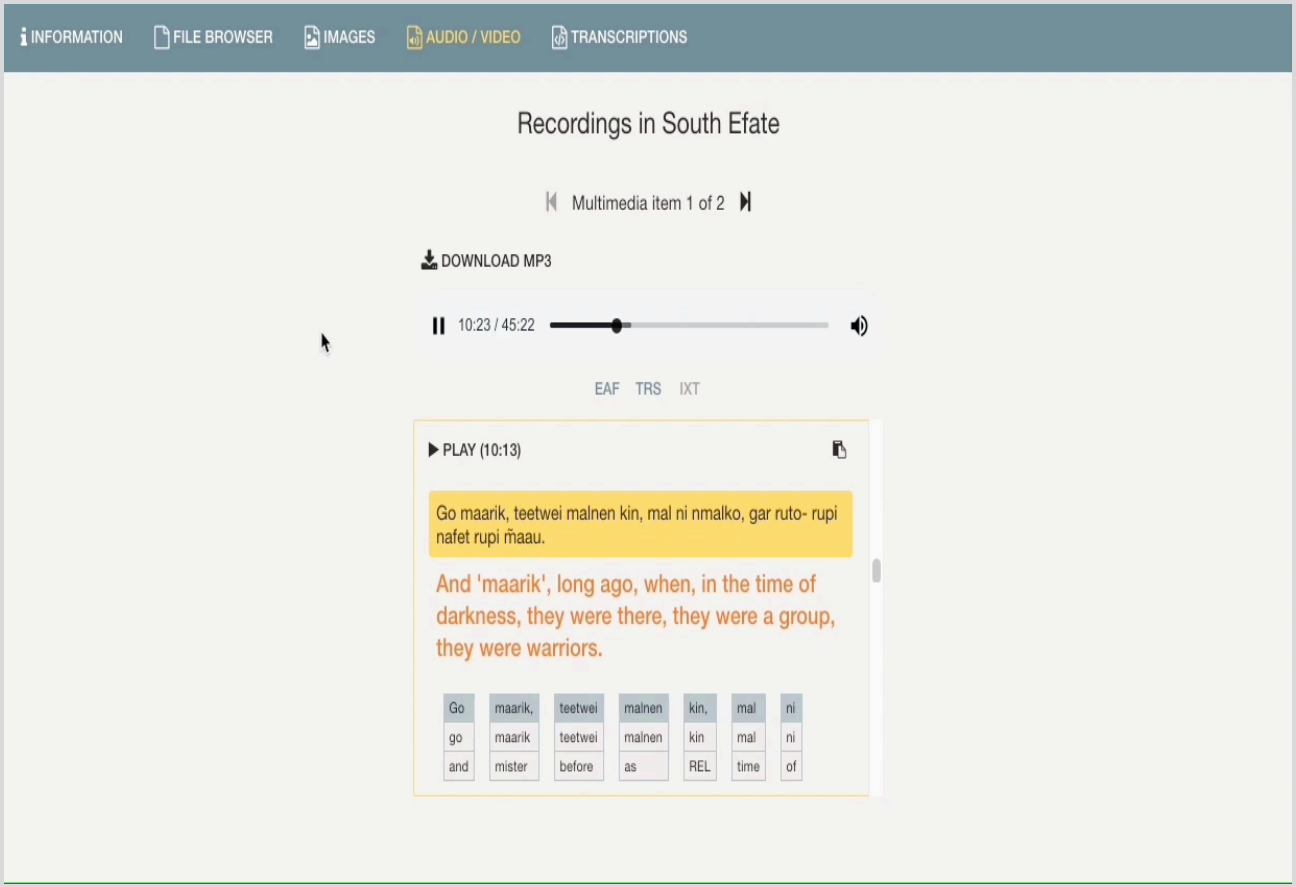
I acknowledge and pay my respect to the Woiwurrung original owners of the land we live and work on

Please sign up:

- https://go.coedl.net/LDocTraining2020

Transcribe





Time-aligned audio and transcript in the PARADISEC archive

Elan

- Transcribe media
- Multiple video files
- Provides a time-aligned transcript of the media
  - Index of the media
  - Useful for subtitling
- Slow down playback
- Automatic insertion of segments based on silences
- Create a corpus of all transcripts to search over
- Allows multiple tiers to be transcribed (speakers, gestures, comments, translations)



Why use Elan?

- Instant access to points in media via text
- Well-known standard format that is being adopted by online services

Installation

- Free download from MPI site
- Version 5.9 (current at 23rd March 2020)
- https://archive.mpi.nl/tla/elan/download
- SimpleElan version 1.3
- Manual updated 2019-10-08

Filetypes

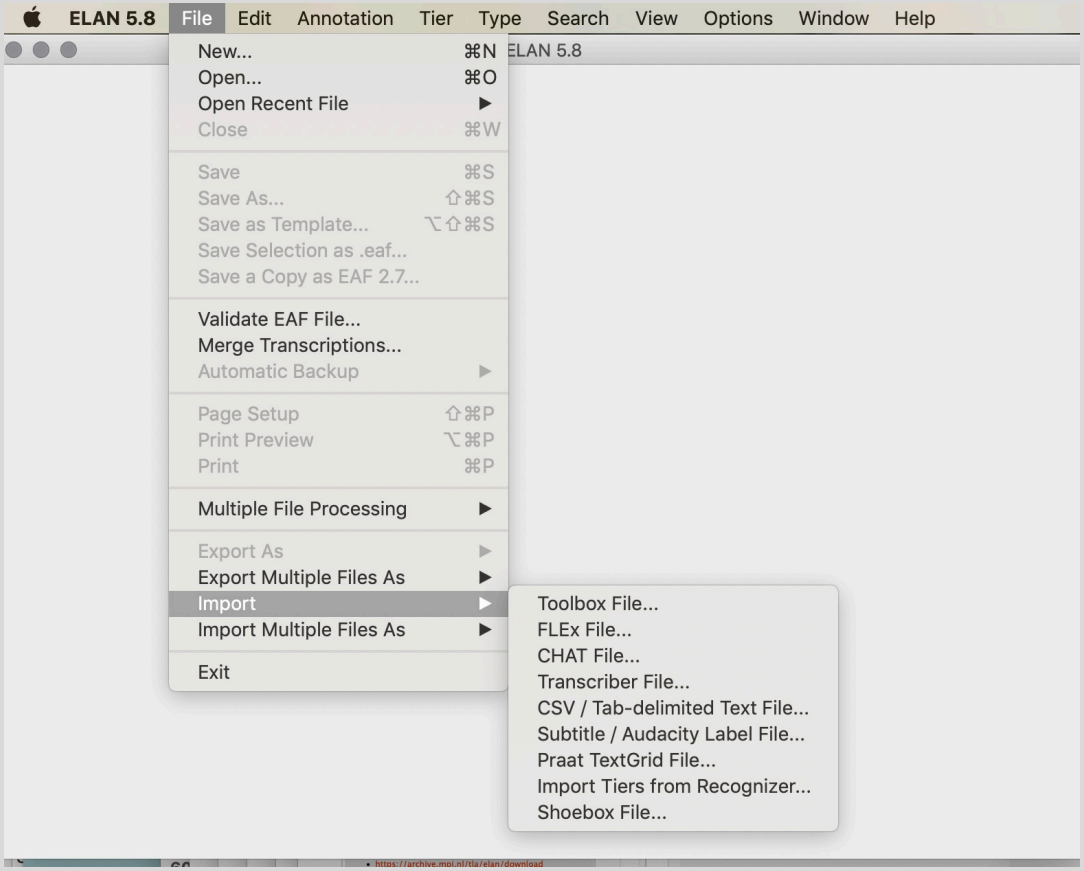
- Whatever your system allows:
- .mov, .avi, .wmv, .mpg
- .wav, .vob, .mp3,

- Annotation is stored in Elan Annotation Format (.eaf)
  - XML, schema-based format (= a good thing)

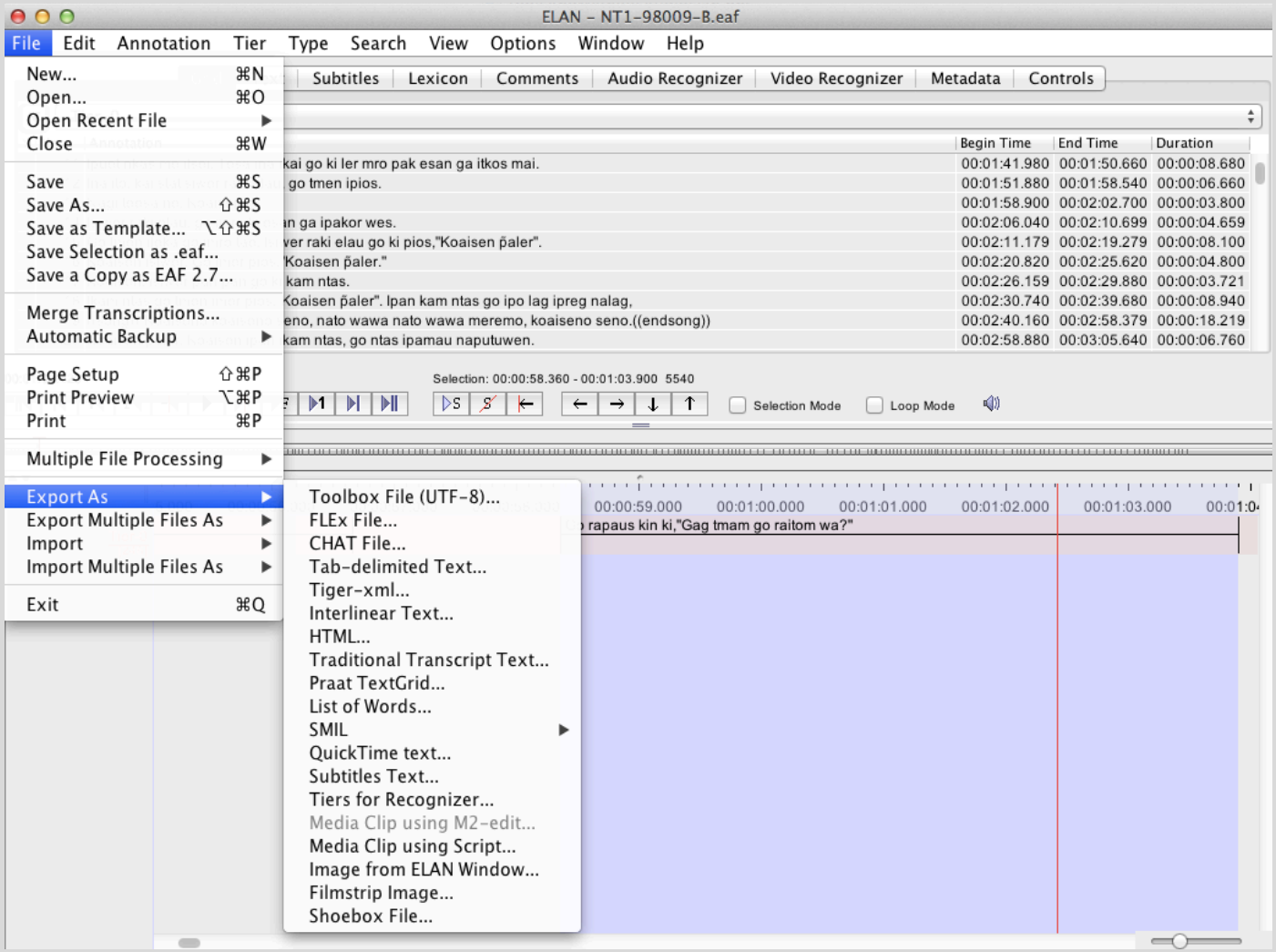
Elan

- To begin with: make an index of what is on a media file
  - could be a summary:
    - first five minutes are general conversation; next ten minutes are singing; next ten minutes is a story told by XXX
  - could be a transcript (word by word typing what is said on the recording)

Elan Imports



Elan Exports



Demo files

provided by MPI

- elan-example1.mpg
- elan-example1.wav
- elan-example1.eaf

- http://www.mpi.nl/tools/elan/elan-example.zip

Simple Elan

- Transcribe media
- No speed controls
- Single tier of transcription
- Allows time-aligning of existing transcript

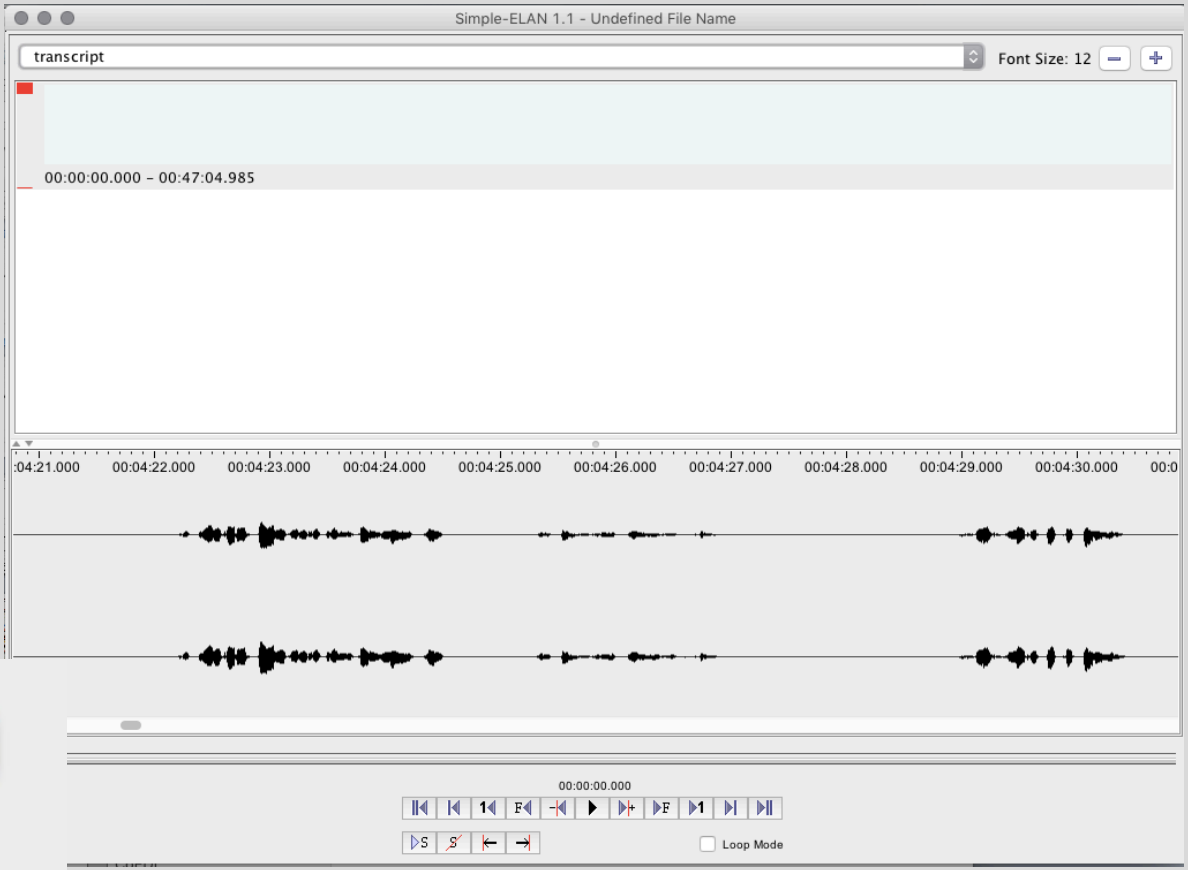


Simple Elan

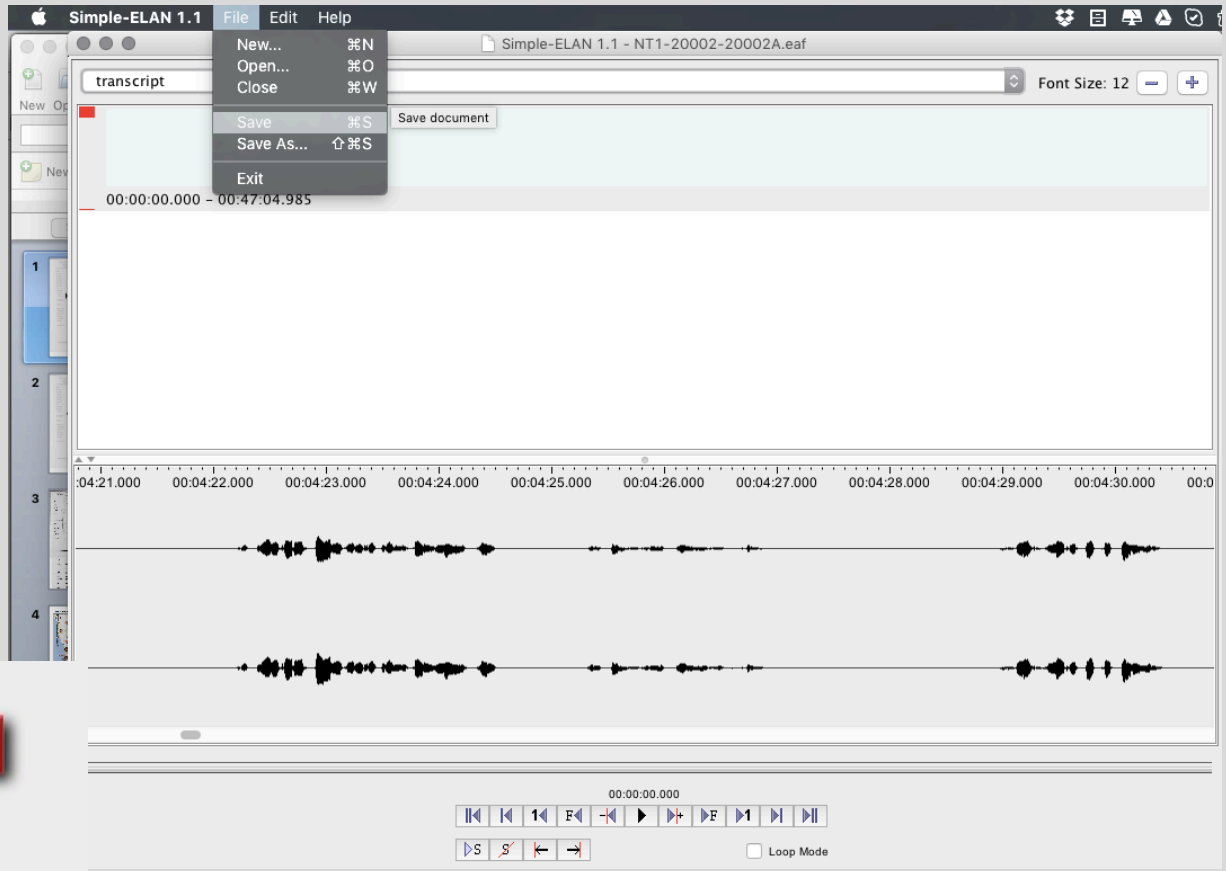
- Allows you to transcribe media
- Drag and drop media file onto the opening screen



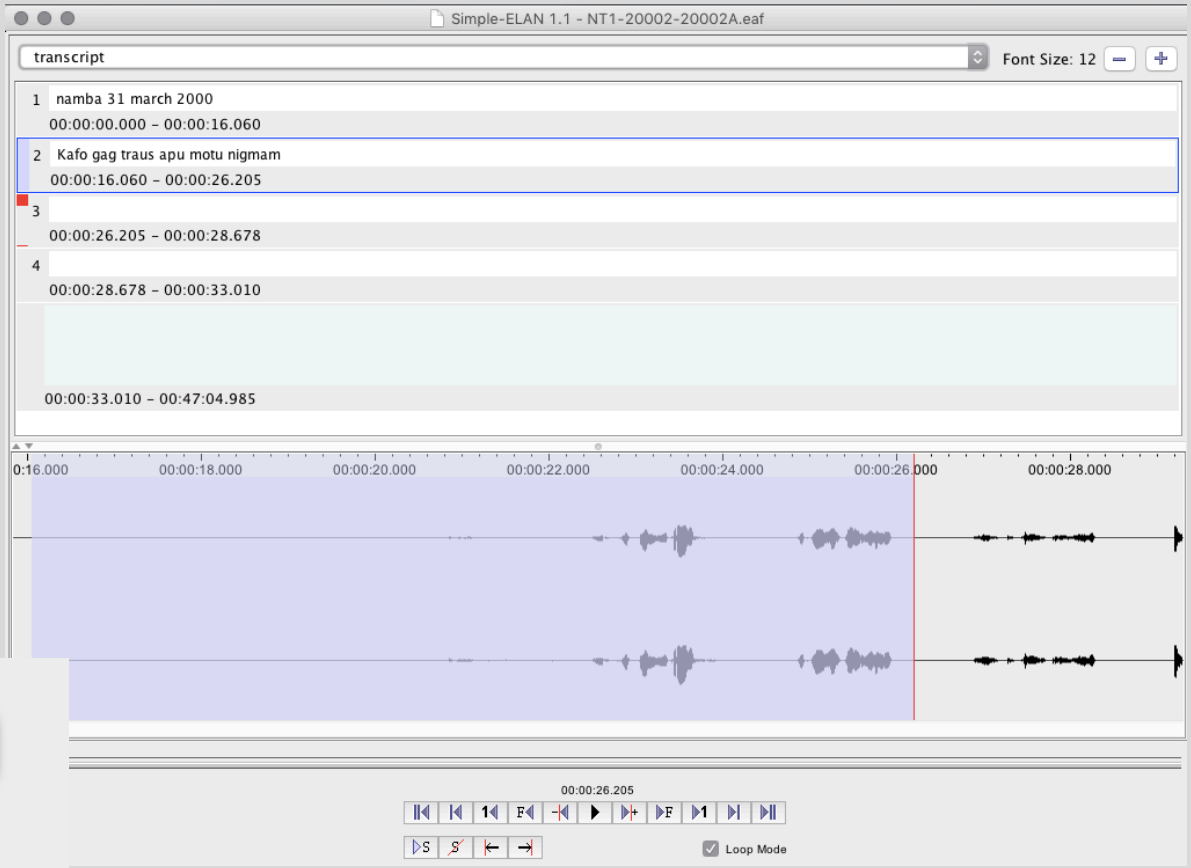
Simple Elan


Simple-ELAN 1.1 - Undefined File Name
transcript
Font Size: 12
00:00:00.000 - 00:47:04.985
:04:21.000
00:04:22.000
00:04:23.000
00:04:24.000
00:04:25.000
00:04:26.000
00:04:27.000
00:04:28.000
00:04:29.000
00:04:30.000
00:0
00:00:00.000
Loop Mode

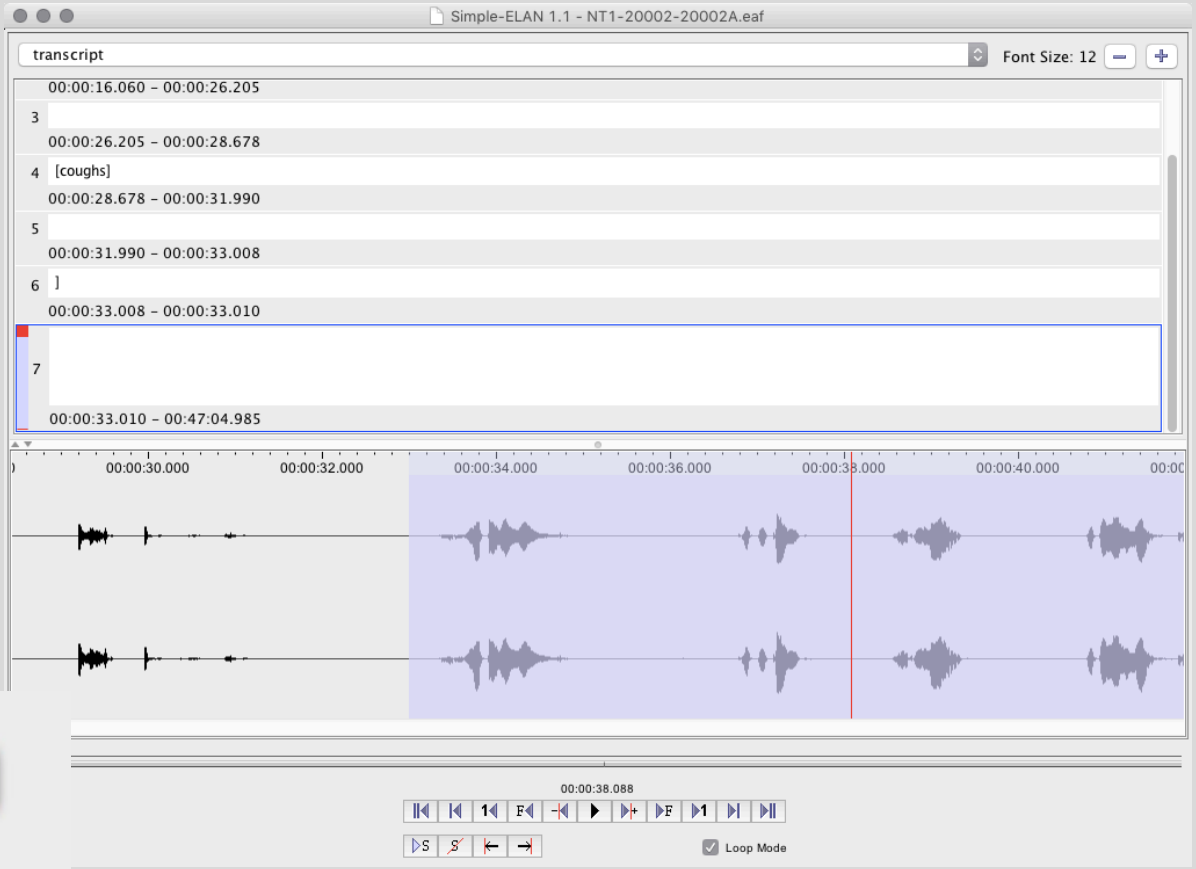
Simple Elan


Simple-ELAN 1.1
File
Edit
Help
New... ⌘N
Open... ⌘O
Close ⌘W
Save ⌘S
Save As... ⇧⌘S
Exit
Simple-ELAN 1.1 - NT1-20002-20002A.eaf
transcript
Font Size: 12
Save document
00:00:00.000 - 00:47:04.985
:04:21.000
00:04:22.000
00:04:23.000
00:04:24.000
00:04:25.000
00:04:26.000
00:04:27.000
00:04:28.000
00:04:29.000
00:04:30.000
00:0
00:00:00.000
Loop Mode

Simple Elan


Simple-ELAN 1.1 - NT1-20002-20002A.eaf
transcript
Font Size: 12
1 namba 31 march 2000
00:00:00.000 – 00:00:16.060
2 Kafo gag traus apu motu nigmam
00:00:16.060 – 00:00:26.205
3
00:00:26.205 – 00:00:28.678
4
00:00:28.678 – 00:00:33.010
00:00:33.010 – 00:47:04.985
0:16.000
00:00:18.000
00:00:20.000
00:00:22.000
00:00:24.000
00:00:26.000
00:00:28.000
00:00:26.205
Loop Mode

Simple Elan


Simple-ELAN 1.1 - NT1-20002-20002A.eaf
transcript
Font Size: 12
00:00:16.060 - 00:00:26.205
3
00:00:26.205 - 00:00:28.678
4 [coughs]
00:00:28.678 - 00:00:31.990
5
00:00:31.990 - 00:00:33.008
6]
00:00:33.008 - 00:00:33.010
7
00:00:33.010 - 00:47:04.985
00:00:30.000
00:00:32.000
00:00:34.000
00:00:36.000
00:00:38.000
00:00:40.000
00:00:38.088
Loop Mode

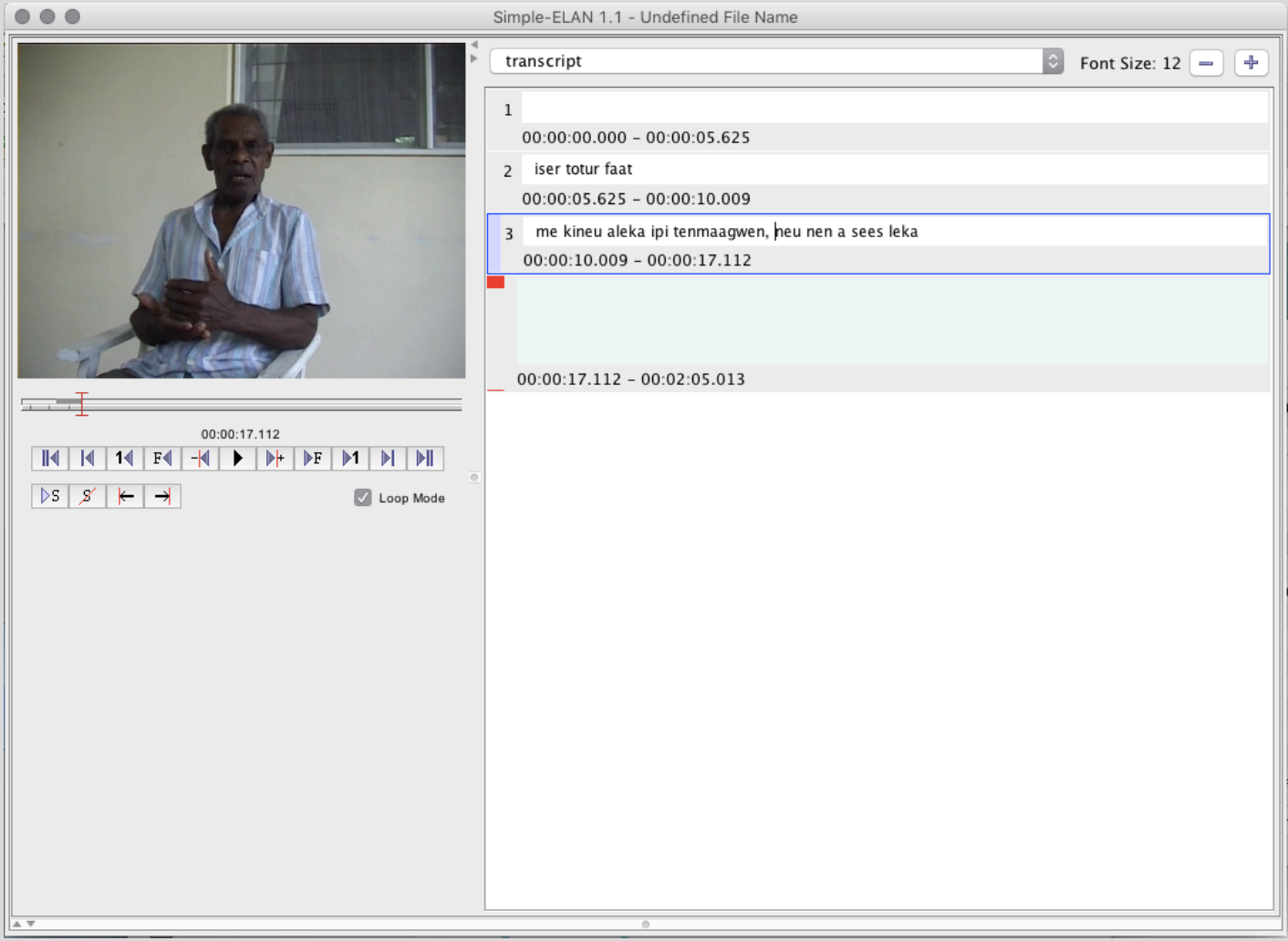
Simple Elan

- gives you a time-aligned index of your media
- needs no setup
- the file it creates can then be opened in Elan



Simple Elan

- Video



Elan sources of information

- https://archive.mpi.nl/tla/elan/documentation
- https://archive.mpi.nl/tla/elan/support
- https://archive.mpi.nl/tla/elan/thirdparty#documentation



Extracting a wav file from a movie

Audacity

https://www.audacityteam.org

Needs ffmpeg extension downloaded

https://lame.buanzo.org/#lameosx64bitdl

TO DOWNLOAD Lame and FFmpeg for Windows, click links BELOW:

The Audacity(r) QA Team suggests users download the ZIP version instead of the .EXE or .DMG (for Mac) versions. If
option, extract the files inside to a well known folder, then open Audacity, go to Library Preferences and configure it to

NOTICE: MacAfee and at least one other antivirus flags this site as unsafe, which is a false positive. FILES I H
www.virustotal.com.

RECOMMENDED Installer Package for Windows: Lame_v3.99.3_for_Windows.exe - (SHA256 SUM here)

ZIP OPTION:
libmp3lame-win-3.99.3.zip (Issues? Some help HERE)

*If you need or want a **newer** version of Lame, because of the **performance improvements with newer AMD and Int***
You can also find 3.98.2 that is known to work with 1.3.x

You can also download the standalone lame v3.100 exe by clicking here - (Virustotal Scan here, dont mind the one fal

---

FFmpeg for Windows and latest Audacity versions
***FFmpeg RECOMMENDED ZIP OPTION:** ffmpeg-win-2.2.2.zip*

FFmpeg 2.2.2 Binary for Windows, compatible with Audacity 2.0.6 and later (please update, or use v0.6.2 belo
ffmpeg-win-2.2.2.exe - (SHA256 SUM here)
(FFmpeg is not for MP3 export for that you need LAME, see above

FFmpeg for older versions of Audacity
FFmpeg 0.6.2 Binary compatible with Audacity 1.3.13 to 2.0.5 only, on Windows:
FFmpeg_v0.6.2_for_Audacity_on_Windows.exe - (ZIP version - here)

FFmpeg and LAME on macOS / Mac OS X

64-bit libraries for Audacity 2.3.1 and later

Audacity 2.3.1 on Mac is now a 64-bit application and thus requires 64-bit libraries for LAME and Ffmpeg.

*It is recommended to use the **.pkg** installer versions (ZIP version is provided as alternative, PKG is better than DMG).*

64-bit LAME for Mac OS X

Mac OSX 64 bit LAME mp3 library for Audacity 2.3.1 and above:

***RECOMMENDED download:**lame_64bit_osx.pkg*

***ZIP Alternative zip download:**Lame_Library_64bit_MAC_OSX_for_Audacity_2.3.1_and_above.zip*

For installation instructions please see LAME installation in the Audacity Manual.

64-bit FFMPEG for Mac OS X

Mac OSX 64 bit FFmpeg library for Audacity 2.3.1 and above:

***RECOMMENDED download:**ffmpeg_64bit_osx.pkg*

***Alternative zip download:**FFmpeg64bit_MAC_OSX_for_Audacity_2.3.1_and_above.zip*

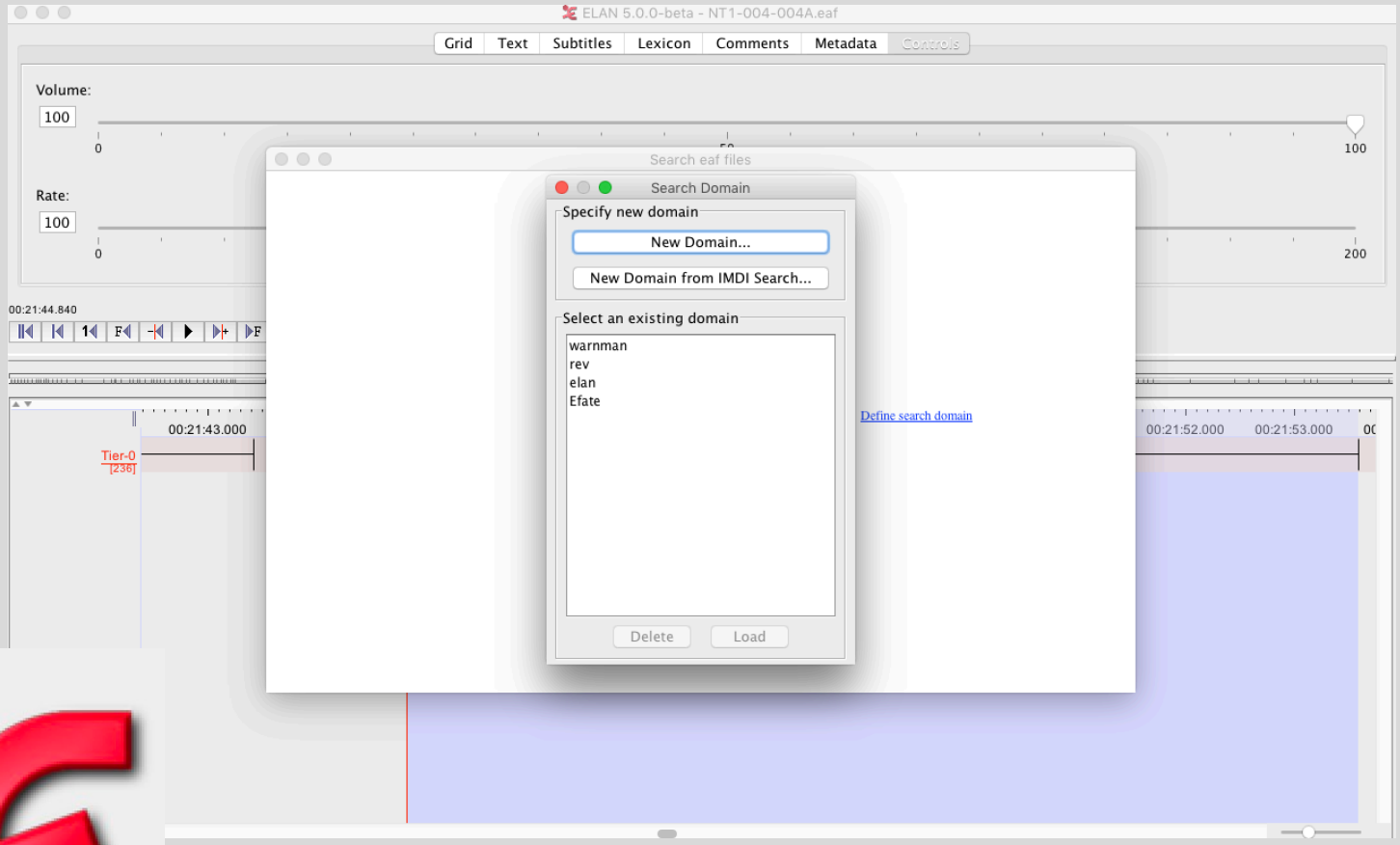
For installation instructions please see FFmpeg installation in the Audacity Manual.

Elan corpus

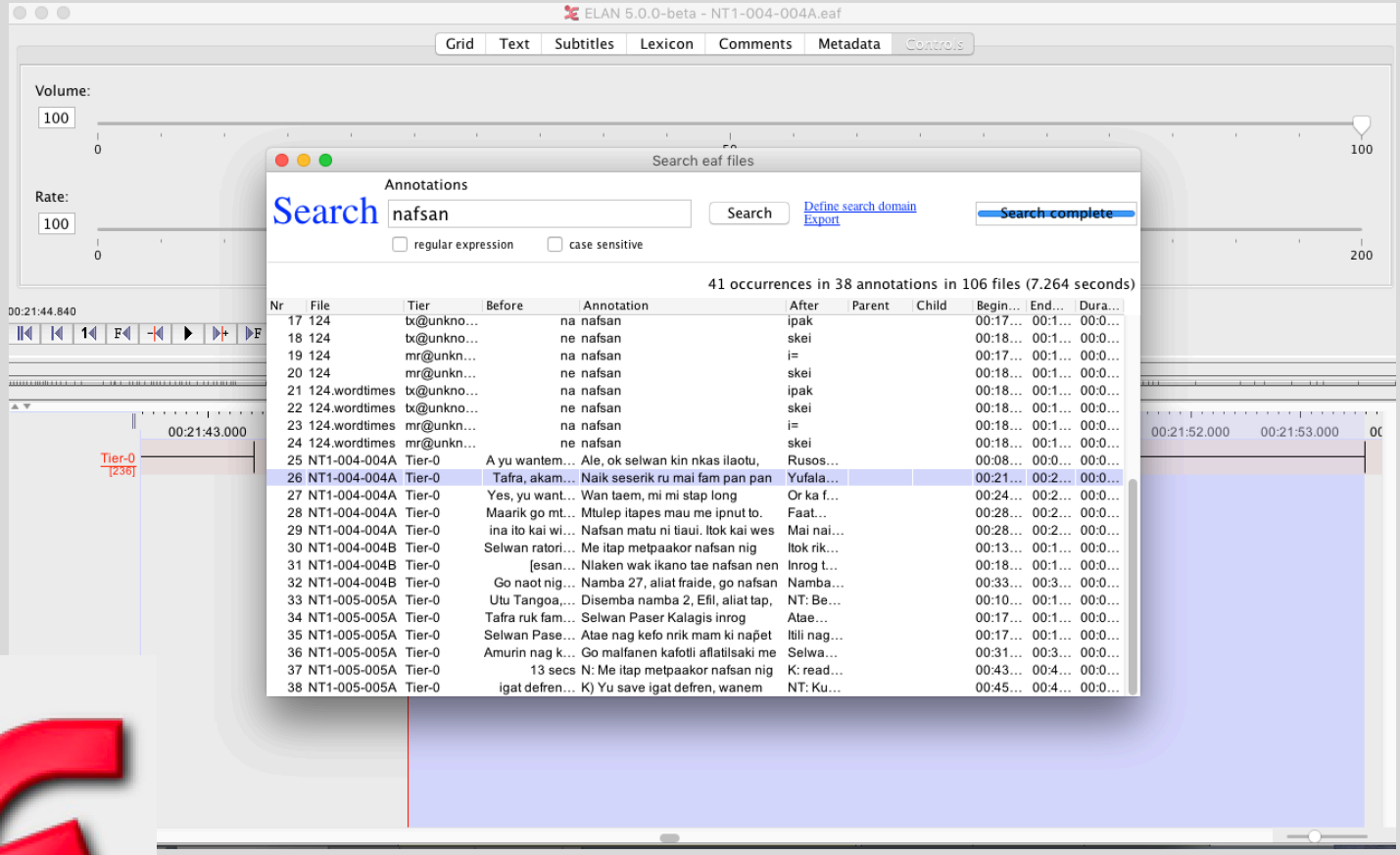
- Search within a document
- Search over a set of Elan files
- Define a folder of files



Elan corpus


ELAN 5.0.0-beta - NT1-004-004A.eaf
Grid
Text
Subtitles
Lexicon
Comments
Metadata
Controls
Volume:
100
0
100
Rate:
100
0
200
00:21:44.840
Search eaf files
Search Domain
Specify new domain
New Domain...
New Domain from IMDI Search...
Select an existing domain
warnman
rev
elan
Efate
Delete
Load
Define search domain
00:21:43.000
00:21:52.000
00:21:53.000
Tier-0
[236]

Elan corpus


ELAN 5.0.0-beta - NT1-004-004A.eaf
Grid
Text
Subtitles
Lexicon
Comments
Metadata
Volume:
100
Rate:
100
Search eaf files
Annotations
Search
nafsan
Search
Define search domain
Export
Search complete
regular expression
case sensitive
41 occurrences in 38 annotations in 106 files (7.264 seconds)
Nr File Tier Before Annotation After Parent Child Begin... End... Dura...
17 124 tx@unkno... na nafsan ipak 00:17... 00:1... 00:0...
18 124 tx@unkno... ne nafsan skei 00:18... 00:1... 00:0...
19 124 mr@unkn... na nafsan i= 00:17... 00:1... 00:0...
20 124 mr@unkn... ne nafsan skei 00:18... 00:1... 00:0...
21 124.wordtimes tx@unkno... na nafsan ipak 00:18... 00:1... 00:0...
22 124.wordtimes tx@unkno... ne nafsan skei 00:18... 00:1... 00:0...
23 124.wordtimes mr@unkn... na nafsan i= 00:18... 00:1... 00:0...
24 124.wordtimes mr@unkn... ne nafsan skei 00:18... 00:1... 00:0...
25 NT1-004-004A Tier-0 A yu wantem... Ale, ok selwan kin nkas ilaotu, Rusos... 00:08... 00:0... 00:0...
26 NT1-004-004A Tier-0 Tafra, akam... Naik seserik ru mai fam pan pan Yufala... 00:21... 00:2... 00:0...
27 NT1-004-004A Tier-0 Yes, yu want... Wan taem, mi mi stap long Or ka f... 00:24... 00:2... 00:0...
28 NT1-004-004A Tier-0 Maarik go mt... Mtulep itapes mau me ipnut to. Faat... 00:28... 00:2... 00:0...
29 NT1-004-004A Tier-0 ina ito kai wi... Nafsan matu ni tiaui. Itok kai wes Mai nai... 00:28... 00:2... 00:0...
30 NT1-004-004B Tier-0 Selwan ratori... Me itap metpaakor nafsan nig Itok rik... 00:13... 00:1... 00:0...
31 NT1-004-004B Tier-0 [esan... Nlaken wak ikano tae nafsan nen Inrog t... 00:18... 00:1... 00:0...
32 NT1-004-004B Tier-0 Go naot nig... Namba 27, aliat fraide, go nafsan Namba... 00:33... 00:3... 00:0...
33 NT1-005-005A Tier-0 Utu Tangoa,... Disemba namba 2, Efil, aliat tap, NT: Be... 00:10... 00:1... 00:0...
34 NT1-005-005A Tier-0 Tafra ruk fam... Selwan Paser Kalagis inrog Atae... 00:17... 00:1... 00:0...
35 NT1-005-005A Tier-0 Selwan Pase... Atae nag kefo nrik mam ki naṕet Itili nag... 00:17... 00:1... 00:0...
36 NT1-005-005A Tier-0 Amurin nag k... Go malfanen kafotli aflatilsaki me Selwa... 00:31... 00:3... 00:0...
37 NT1-005-005A Tier-0 13 secs N: Me itap metpaakor nafsan nig K: read... 00:43... 00:4... 00:0...
38 NT1-005-005A Tier-0 igat defren... K) Yu save igat defren, wanem NT: Ku... 00:45... 00:4... 00:0...
00:21:43.000
00:21:52.000
00:21:53.000
Tier-0

Starting Elan

- New file
- Select media (up to 4 media files can be jointly annotated)
- You can choose a ‘template’ that has existing settings
- Name media and annotation file the same (e.g fileX.mpg, fileX.eaf)

Elan

- Save your file right away: File>Save. The ELAN file extension is .eaf.
- Set your Automatic backup to 1 or 5 minutes. File>Automatic Backup>5 minutes.



- To add an annotation, double-click inside the blue area. Type your annotation.

- To delete an annotation, click the annotation so that the horizontal line is blue. Then use Edit>Delete Annotation.

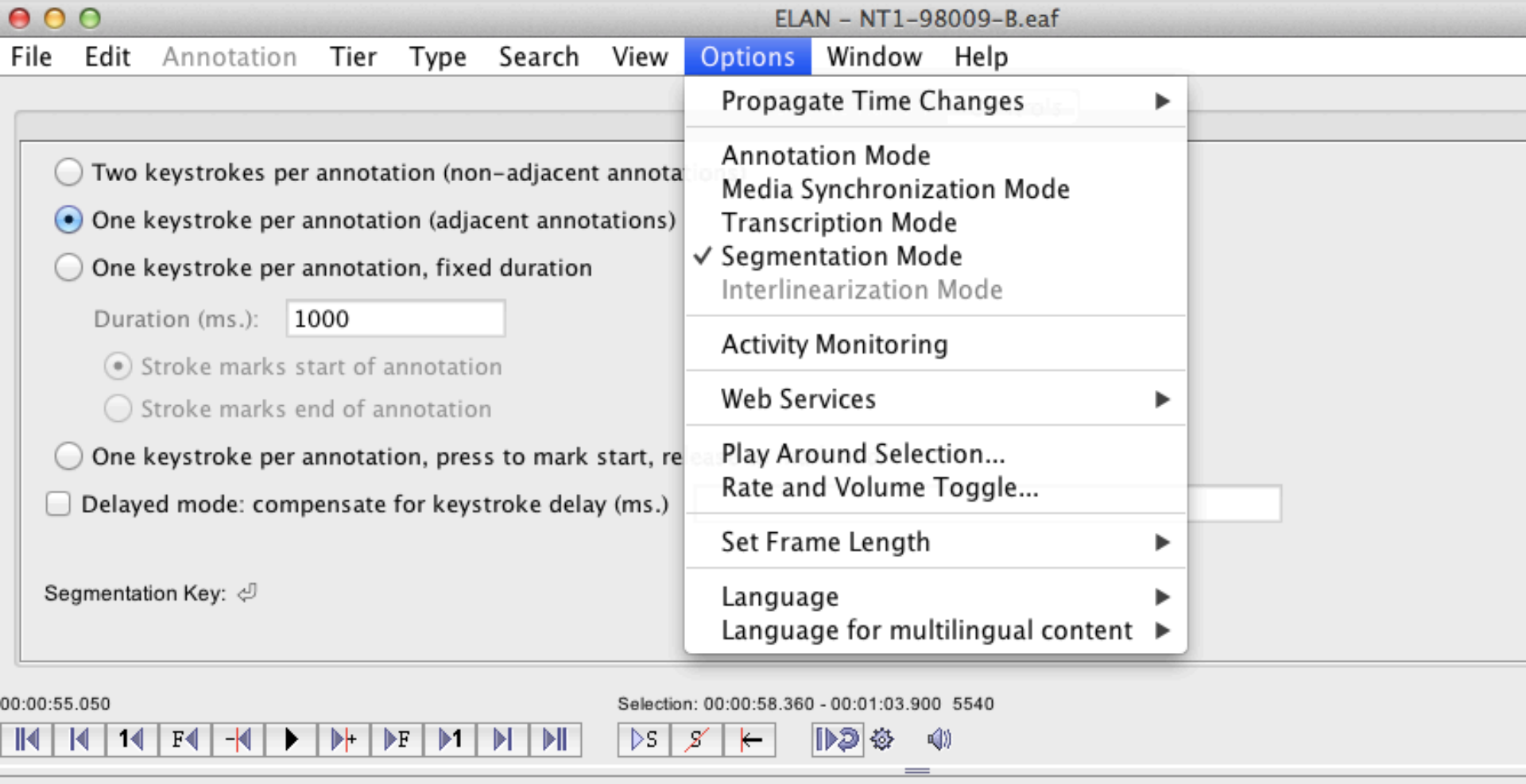


Variable playback speed

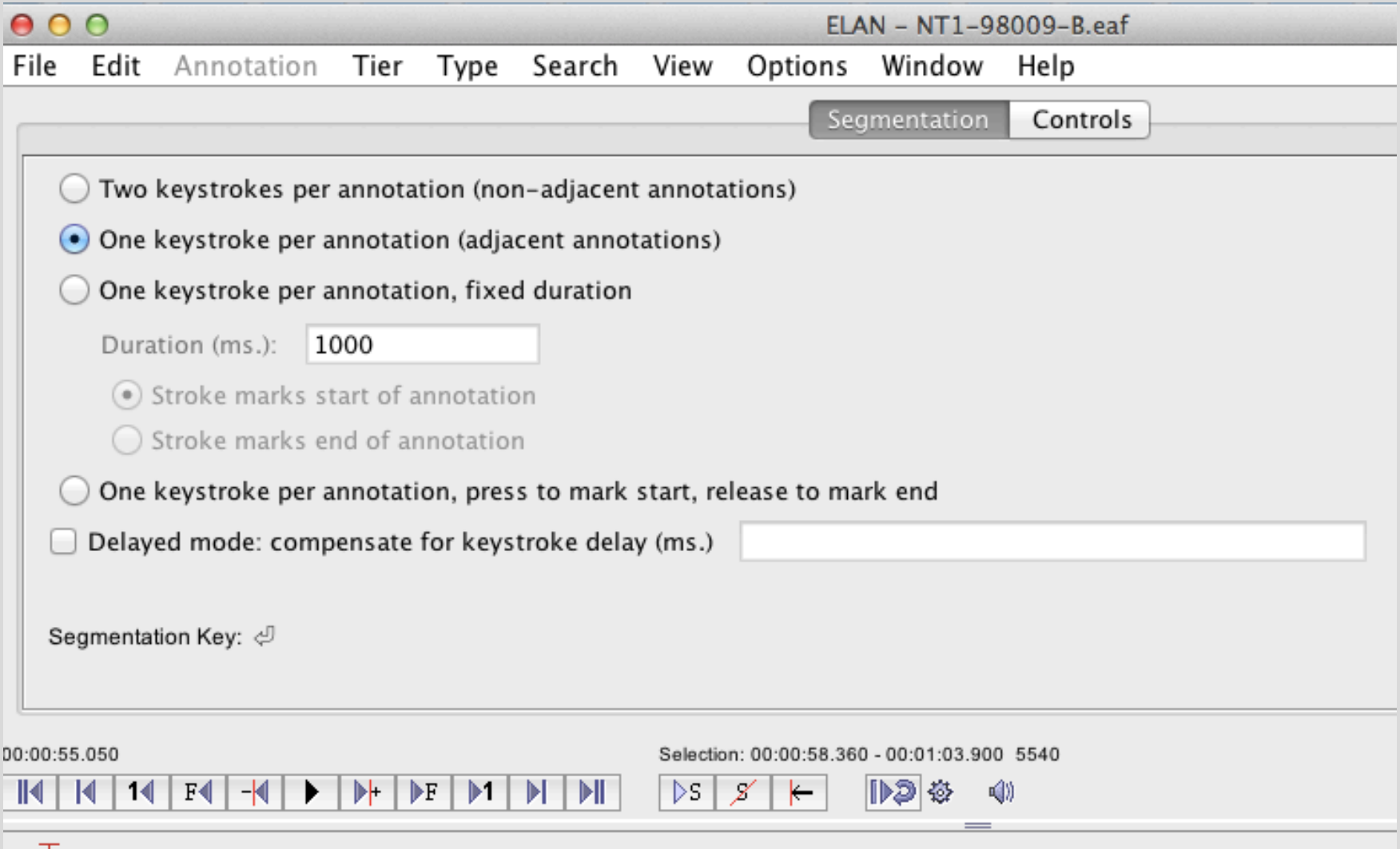
- Controls – Rate
- Loop – set a ‘chunk’ then loop over it until you transcribe it



Quick segmentation



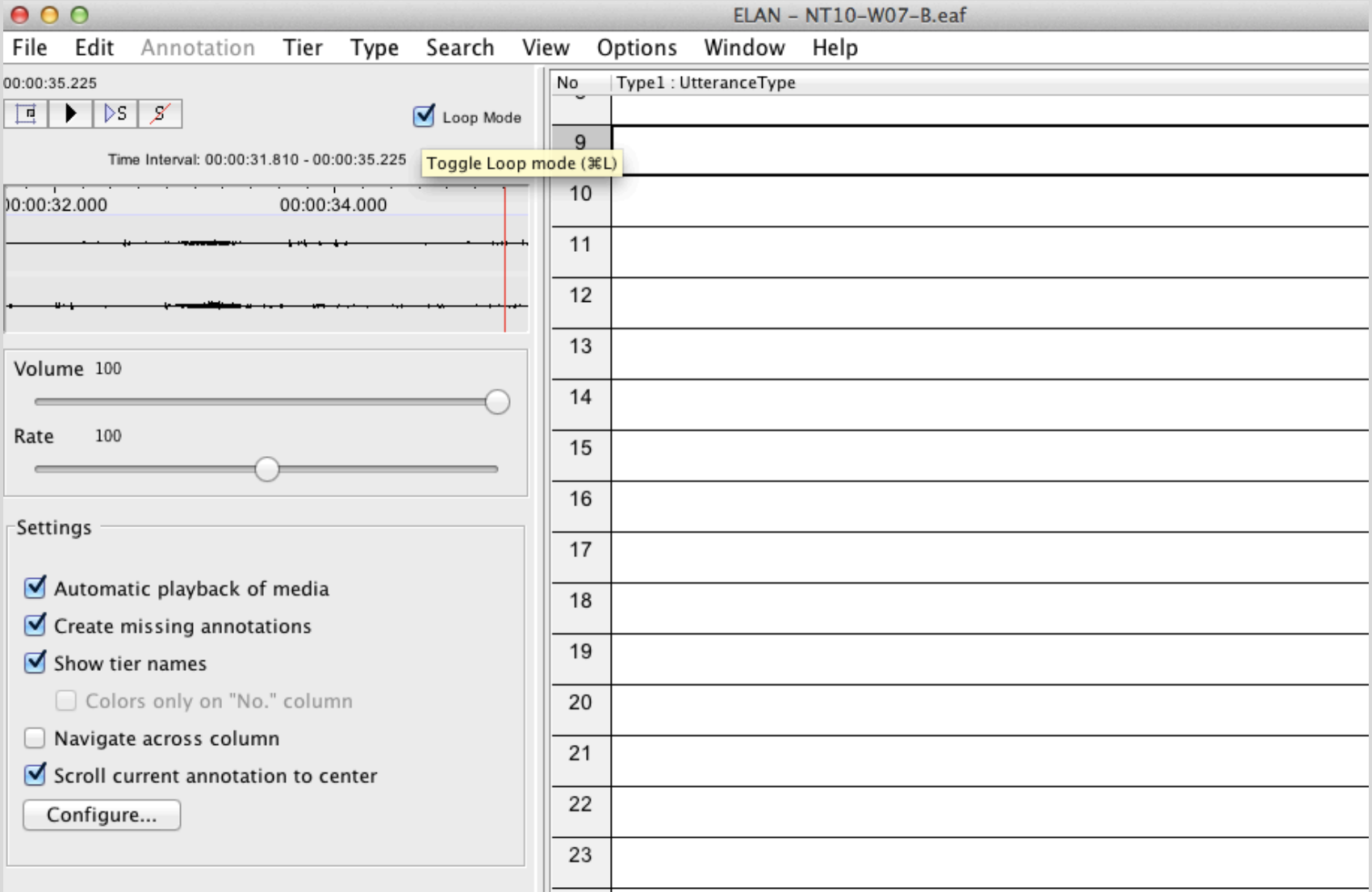
Quick segmentation



Quick segmentation

- This results in an Elan file that has chunks or segments ready to be typed into
- Saves lots of time
- Set ‘loop’ then enter transcript

Transcription mode



Audio recognizer

Silence Analyzer

- Chunks the media according to ‘silence’
- You can determine what silence is and how long it should be

