

Introduction to Elan

Nick Thieberger

School of Languages and Linguistics

The University of Melbourne

ARC Centre of Excellence for the Dynamics of Language



Transcribe

Go RU PES ICI NAFAN, TETE NEN KIN ITHA NAFAN.
Mo di tak tak long mund. Some we oli safe tak tak
INRUS ITHA PES, KE FO ITHA PES, GO AMRO SOK NAFAN
oli tak tak sand, emi tak tak Mo mi Remember wan
WEN NI MARIK KICEI RU NEU ~~RENGI~~, TIL USUS
Storv blang wan olfala. Oli nekam blang mi. ~~oli emi tak tak~~
NAFET, TETE MARIK NI NAFAR GAKIT NI EFIL,
olgeta, Saui olfala blang ITHA Island,
RU MAI RU ITHA NAFAR GAKIT NI EFAT NE,
oli kam oli nitim ol man blang you mi long Efata.
RU PI TENI NAUR EFIL, GO KA FO TIL NAFAR
olgeta oli blang Fila Island. So bambae ni inlem nem
TEFAN, SKEI IPI MARIK NAFAR GO ISKEI, MARIK
blang olgeta olsem, wan emi olfala NAFAR, mo wan em olfala
KORU, GO ISKEI IPI MARIK ~~PARO~~ KOTO, KEO ISKEI
KORI, mo wan emi olfala ~~Boko~~ KOTO, mo wan emi
ipi MARIK ~~MARATAMAM~~, NIE RU NA GAKIT ISKEI
olfala ~~MARATAMAM~~, Be wan long olgeta nem blang
NAGIEN IPI MARIK, ANIET PAKOR NAGIEN, GA ITHA
olfala mi forgetem, nem blang em, Emi safe
PES INRUS ITHA PES ICI NAFAN NI, NAFAN NI ETOMG
tak tak, Emi safe tak tak sand long lamiis, lamiis blang
wait man.

GO GAKIT IPI NAFANOL NEN ITHA GAKIT NRE NAFAN
So emi now emi man we emi stop tanaw tak tak blang
-NI ~~EDOMG~~ ITHA PAK NAFAN GAKIT, GO INI TAWI
ol wait man ikama long, lamiis blang you mi. Mo emi nekam
NEN RU THU, PE ITHA TEFAN KI. GO RU PAN RU
Sambing long ol olfala em, Emi talam olsem, So oli ko oli Salaw
SOR NAFAN, ICI RU SORI NI TI TIKAZ E MATH NI, KIN
ground. Oli Salaw ikam kam, E tan blaw tedes
NIB, ITHA KIN TUNA EFIL ^{NAUR} LAMATE, EIAL PORT VILATAM
you mi stop talam FICA Island, VILA, PORT VILATAM
ITHAS, GO NABI MARIK NE, IMAI, RU SOR NAFAN,
istap long em, Mo nem blang olfala ia, ikam oli Salaw ground
MLAKEM TETERARU NEN RU MAI RU PAN TOTE
From Sam ship we oli kam, oli kam oli ko Ankar
WATH EFIL, RU PAMORI NA IPI NAOR NI GO RU
long Faw VILA, Oli tak se emi wan good fala pes,
MLIR NA RUK FAK TOTE, GO RU SORI TU NAFET
So oli wantem pem So oli Salaw iko long
TAWI NE RU TAWI SORI, RU SORI TU TE TAWI TAWI NE
Sam, oli Salaw iko long sam olfala wait
NATAMOL ITHA NE, GO IPAKTOFI, TETE NATAMOL RU PAN
man in So emi pem, Some fala man oli ko
PES MARIK KICEI KIA RU RU SO SO KI, MARIK SIK SIK
Eke tak long wan olfala, emi oli Sigantem, olfala sal sal

Me, selwan ag kupi eñae, tiawi itraus
traus traus traus, natrauswen ga itaos
nlag. Itrausi pan kaipa. Me komam uta laap
kin uto mau, a? Malen umat, inom.

*But when you are far away the old man can
talk and talk and talk, his story is like
the wind. He tells it and it is gone. But
there aren't many of us left. When we die,
it will be finished.*



(NT1-98009-A, 1932.36 1942.16)

†Kalfañun Mailei, 1998, Erakor Village, Efate, Vanuatu

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

Recordings in South Efate

Multimedia item 1 of 2

DOWNLOAD MP3

10:23 / 45:22

EFATRSIXT

PLAY (10:13)

Go maarik, teetwei mainen kin, mal ni nmalko, gar ruto- rupi nafet rupi maau.

And 'maarik', long ago, when, in the time of darkness, they were there, they were a group, they were warriors.

Go	maarik,	teetwei	mainen	kin,	mal	ni
go	maarik	teetwei	mainen	kin	mal	ni
and	mister	before	as	REL	time	of

Time – aligned audio and transcript in the PARADISEC archive

Simple Elan

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

- 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)

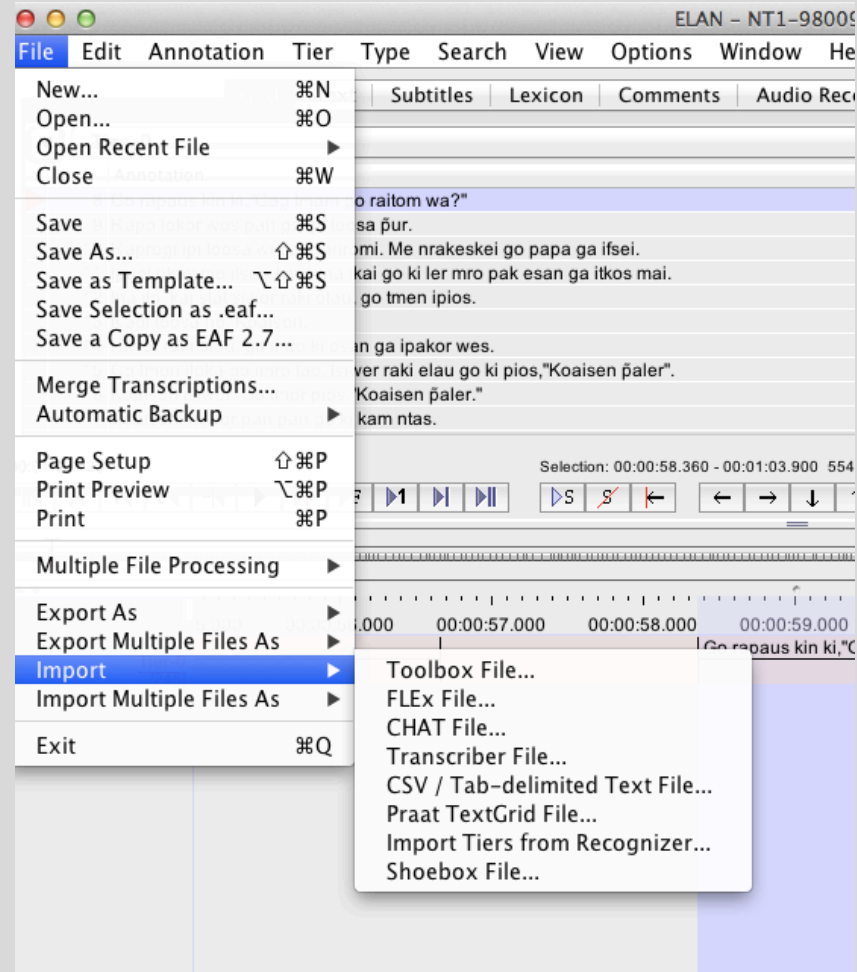
Installation

- Free download from MPI site
- Version 5.7 (current at July 2019)
- <http://tla.mpi.nl/tools/tla-tools/elan/>
- SimpleElan version 1.1
- Manual updated 2018-12-05

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 Imports



Elan Exports

The screenshot displays the ELAN software interface with the 'Export As' menu open. The window title is 'ELAN - NT1-98009-B.eaf'. The menu options are as follows:

- File
- Edit
- Annotation
- Tier
- Type
- Search
- View
- Options
- Window
- Help

Submenus include: Subtitles, Lexicon, Comments, Audio Recognizer, Video Recognizer, Metadata, and Controls.

	Begin Time	End Time	Duration
kai go ki ler mro pak esan ga itkos mai.	00:01:41.980	00:01:50.660	00:00:08.680
go tmen ipios.	00:01:51.880	00:01:58.540	00:00:06.660
	00:01:58.900	00:02:02.700	00:00:03.800
in ga ipakor wes.	00:02:06.040	00:02:10.699	00:00:04.659
ver raki elau go ki pios,"Koaisen paler".	00:02:11.179	00:02:19.279	00:00:08.100
Koaisen paler."	00:02:20.820	00:02:25.620	00:00:04.800
kam ntas.	00:02:26.159	00:02:29.880	00:00:03.721
Koaisen paler". Ipan kam ntas go ipo lag ipreg nalag,	00:02:30.740	00:02:39.680	00:00:08.940
eno, nato wawa nato wawa meremo, koaiseno seno.(endsong)	00:02:40.160	00:02:58.379	00:00:18.219
kam ntas, go ntas ipamau naputuwen.	00:02:58.880	00:03:05.640	00:00:06.760

Selection: 00:00:58.360 - 00:01:03.900 5540

Export As menu options:

- Toolbox File (UTF-8)...
- FLEX File...
- CHAT File...
- Tab-delimited Text...
- Tiger-xml...
- Interlinear Text...
- HTML...
- Traditional Transcript Text...
- Praat TextGrid...
- List of Words...
- SMIL
- QuickTime text...
- Subtitles Text...
- Tiers for Recognizer...
- Media Clip using M2-edit...
- Media Clip using Script...
- Image from ELAN Window...
- Filmstrip Image...
- Shoobox File...

Demo files

provided by MPI

- elan-example1.mpg
- elan-example1.wav
- elan-example1.eaf
- <http://www.mpi.nl/tools/elan/elan-example.zip>

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)

Simple Elan

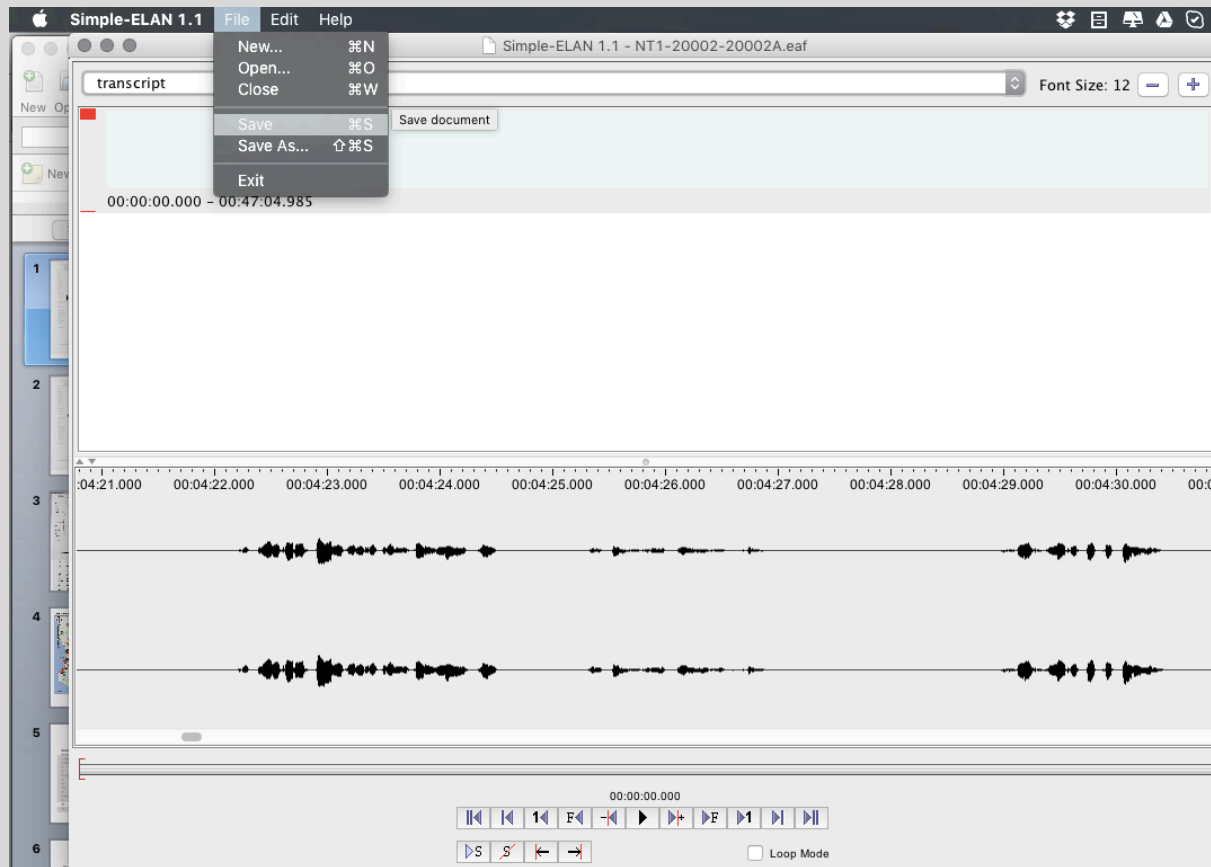
- Allows you to transcribe media
- Drag and drop media file onto the opening screen



Simple Elan

The screenshot displays the Simple-ELAN 1.1 software interface. At the top, the window title is "Simple-ELAN 1.1 - Undefined File Name". Below the title bar is a search bar containing the text "transcript" and a font size control set to "Font Size: 12". The main area is divided into two sections: a large white text area for the transcript and a lower section for audio waveforms. The transcript area shows a time range of "00:00:00.000 - 00:47:04.985". The audio waveforms are displayed on a timeline with time markers from 00:04:21.000 to 00:04:30.000. The interface includes a playback control bar at the bottom with various buttons for navigation and a "Loop Mode" checkbox.

Simple Elan



Simple Elan

The screenshot displays the Simple-ELAN 1.1 software interface. The window title is "Simple-ELAN 1.1 - NT1-20002-20002A.eaf". The main area is divided into two sections: a transcript editor at the top and an audio waveform at the bottom.

The transcript editor shows a list of segments:

- 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

The second segment is highlighted with a blue border. Below the transcript, the audio waveform is displayed, showing the amplitude of the audio signal over time. The time axis ranges from 00:00:16.000 to 00:00:28.000. A red vertical line is positioned at 00:00:26.205, corresponding to the end of the second segment.

At the bottom of the interface, there is a playback control bar with various buttons for navigation and playback. The current time is displayed as 00:00:26.205. A "Loop Mode" checkbox is checked.

Simple Elan

The screenshot displays the Simple-ELAN 1.1 software interface. The window title is "Simple-ELAN 1.1 - NT1-20002-20002A.eaf". The interface is divided into two main sections: a transcript editor and an audio waveform viewer.

Transcript Editor: The transcript is titled "transcript" and has a font size of 12. It contains seven numbered segments with their corresponding time intervals:

- 3: 00:00:16.060 - 00:00:26.205
- 4: [coughs] 00:00:26.205 - 00:00:28.678
- 5: 00:00:28.678 - 00:00:31.990
- 6:] 00:00:31.990 - 00:00:33.008
- 7: 00:00:33.008 - 00:00:33.010
- 00:00:33.010 - 00:47:04.985

The seventh segment is currently selected, indicated by a blue border. The transcript is displayed in a scrollable area.

Audio Waveform Viewer: The waveform viewer shows two channels of audio data. The time axis is labeled from 00:00:30.000 to 00:00:40.000. A red vertical line indicates the current playback position at 00:00:38.088. The waveform shows several distinct peaks, corresponding to the segments in the transcript.

Playback Controls: At the bottom of the interface, there is a set of playback controls including buttons for play/pause, stop, previous, next, and full screen. A "Loop Mode" checkbox is checked.

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

The screenshot displays the Simple-ELAN 1.1 software interface. The window title is "Simple-ELAN 1.1 - Undefined File Name". The interface is divided into several sections:

- Video Player:** Located on the left, it shows a video of an elderly man with grey hair and glasses, wearing a light blue and white striped short-sleeved shirt, sitting in a white chair and speaking. Below the video is a progress bar and a control panel with various playback buttons (play, stop, previous, next, full screen, etc.) and a "Loop Mode" checkbox which is checked.
- Transcript Window:** Located on the right, it has a search bar containing the word "transcript" and a "Font Size: 12" setting. The transcript is organized into numbered segments:
 - Segment 1: 00:00:00.000 - 00:00:05.625
 - Segment 2: 00:00:05.625 - 00:00:10.009, containing the text "iser totur faat".
 - Segment 3: 00:00:10.009 - 00:00:17.112, containing the text "me kineu aleka ipi tenmaagwen, heu nen a sees leka". This segment is currently selected with a blue border.
 - Segment 4: 00:00:17.112 - 00:02:05.013

Elan

Use template

A set of ELAN templates for free download.

There are six templates covering the following types of requirements:

- Template1.etf: 1 speaker, tiers for Text and Free Translation.
- Template2.etf: 2 speakers, tiers for Text and Free Translation.
- Template3.etf: 2 speakers, tiers for Text, Free Translation, and Gesture.
- Template4.etf: 1 speaker, tiers for Text, Word, Gloss, and Free Translation.
- Template5.etf: 2 speakers, tiers for Text, Word, Gloss, and Free Translation.
- Template6.etf: 2 speakers, tiers for Text, Word, Gloss, Free Translation, and Gesture.

- Save your file right away: File>Save. The ELAN file extension is .eaf.
- Set your Automatic backup to 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.

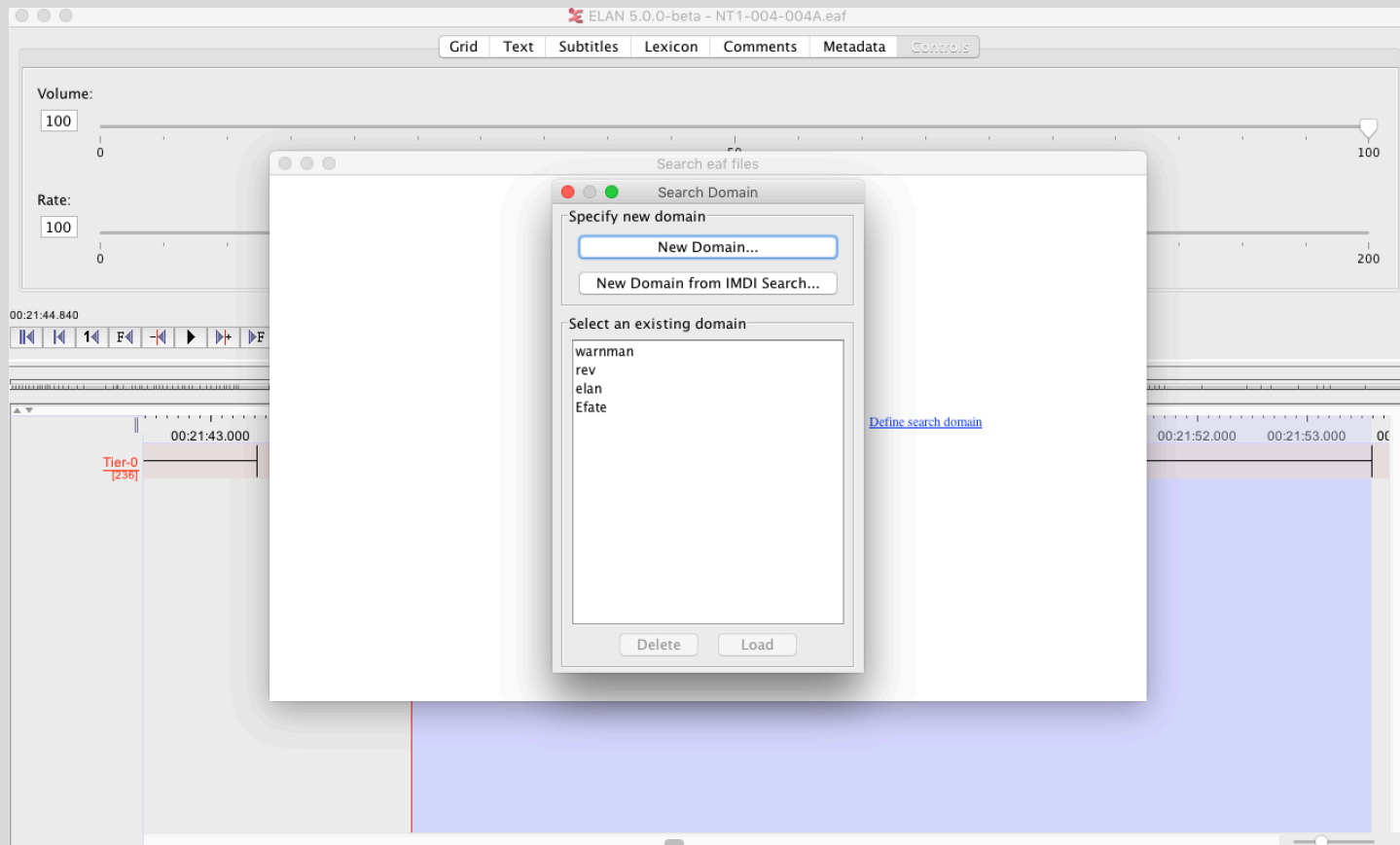
Variable playback speed

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

Elan corpus

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

Elan corpus



Elan corpus

ELAN 5.0.0-beta - NT1-004-004A.eaf

Grid Text Subtitles Lexicon Comments Metadata Controls

Volume: 100

Rate: 100

Search eaf files

Annotations

Search 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	lx@unkno...	na	nafsan	ipak			00:17...	00:1...	00:0...
18	124	lx@unkno...	ne	nafsan	skei			00:18...	00:1...	00:0...
19	124	mr@unkn...	na	nafsan	=			00:17...	00:1...	00:0...
20	124	mr@unkn...	ne	nafsan	skei			00:18...	00:1...	00:0...
21	124.wordtimes	lx@unkno...	na	nafsan	ipak			00:18...	00:1...	00:0...
22	124.wordtimes	lx@unkno...	ne	nafsan	skei			00:18...	00:1...	00:0...
23	124.wordtimes	mr@unkn...	na	nafsan	=			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...	Niaken wak ikano tae nafsan nen	Inrog t...			00:18...	00:1...	00:0...
32	NT1-004-004B	Tier-0	Go naot nig...	Namba 27, alat fraide, go nafsan	Namba...			00:33...	00:3...	00:0...
33	NT1-005-005A	Tier-0	Utu Tangoa...	Disemba namba 2, Efil, alat 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 napet	Itili nag...			00:17...	00:1...	00:0...
36	NT1-005-005A	Tier-0	Amurin nag k...	Go malfanen kafotli aflatlisaki 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...

Quick segmentation

The screenshot displays the ELAN software interface for the file "ELAN - NT1-98009-B.eaf". The menu bar includes "File", "Edit", "Annotation", "Tier", "Type", "Search", "View", "Options", "Window", and "Help". The "Options" menu is open, showing the following items: "Propagate Time Changes", "Annotation Mode", "Media Synchronization Mode", "Transcription Mode", "Segmentation Mode" (checked), "Interlinearization Mode", "Activity Monitoring", "Web Services", "Play Around Selection...", "Rate and Volume Toggle...", "Set Frame Length", "Language", and "Language for multilingual content".

In the background, the "Options" dialog box is visible, with the "Segmentation Mode" radio button selected. Other options include "Two keystrokes per annotation (non-adjacent annotations)", "One keystroke per annotation (adjacent annotations)", "One keystroke per annotation, fixed duration" (with a "Duration (ms.):" field set to "1000"), "Stroke marks start of annotation", "Stroke marks end of annotation", "One keystroke per annotation, press to mark start, release to mark end", and "Delayed mode: compensate for keystroke delay (ms.)". The "Segmentation Key" is set to the left arrow key.

At the bottom of the window, the playback controls show a time of "00:00:55.050" and a selection range of "00:00:58.360 - 00:01:03.900 5540". The playback buttons include play/pause, stop, previous, next, and various navigation controls.

Quick segmentation

The screenshot shows the ELAN software interface with the 'Segmentation' panel active. The window title is 'ELAN - NT1-98009-B.eaf'. The menu bar includes 'File', 'Edit', 'Annotation', 'Tier', 'Type', 'Search', 'View', 'Options', 'Window', and 'Help'. The 'Segmentation' panel has two tabs: 'Segmentation' (selected) and 'Controls'. The 'Segmentation' panel contains the following options:

- Two keystrokes per annotation (non-adjacent annotations)
- One keystroke per annotation (adjacent annotations)
- One keystroke per annotation, fixed duration
 - Duration (ms.):
 - 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: ↵

At the bottom of the window, the time display shows '00:00:55.050' and 'Selection: 00:00:58.360 - 00:01:03.900 5540'. The playback controls include buttons for play/pause, stop, previous, first, next, last, and a volume icon.

Quick segmentation

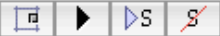
- This results in an Elan file that has chunks or segments ready to be typed into
- Saves lots of time
- Set 'loop' then enter transcript

Transcription mode

ELAN - NT10-W07-B.eaf

File Edit Annotation Tier Type Search View Options Window Help

00:00:35.225

 Loop Mode

Time Interval: 00:00:31.810 - 00:00:35.225 **Toggle Loop mode (⌘L)**

00:00:32.000 00:00:34.000

Volume 100

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

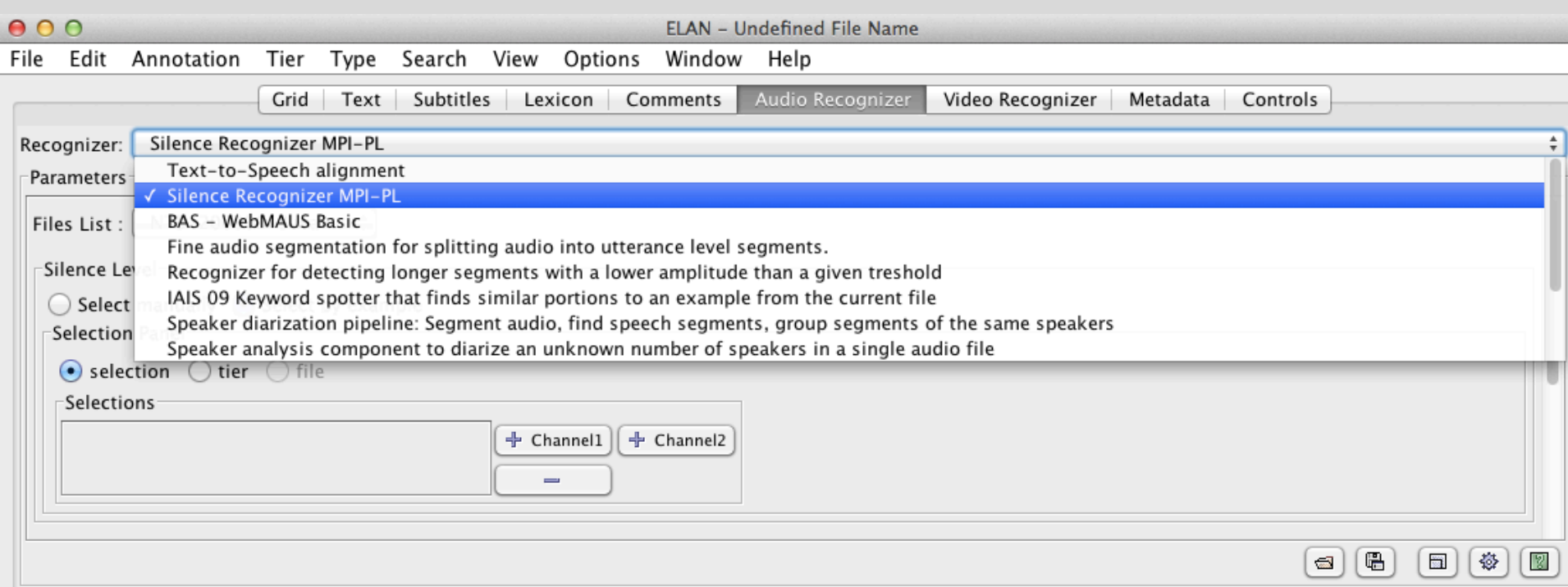
Configure...

No.	Type1 : UtteranceType
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	

Audio recognizer

Silence Analyzer

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



Elan sources of information

- “Using Elan” (Albert Bickford)
- Elan Manual (Birgit Hellwig et al)
- Andrea Berez’s course on Elan
- Elan mailing list
- CoLang courses