

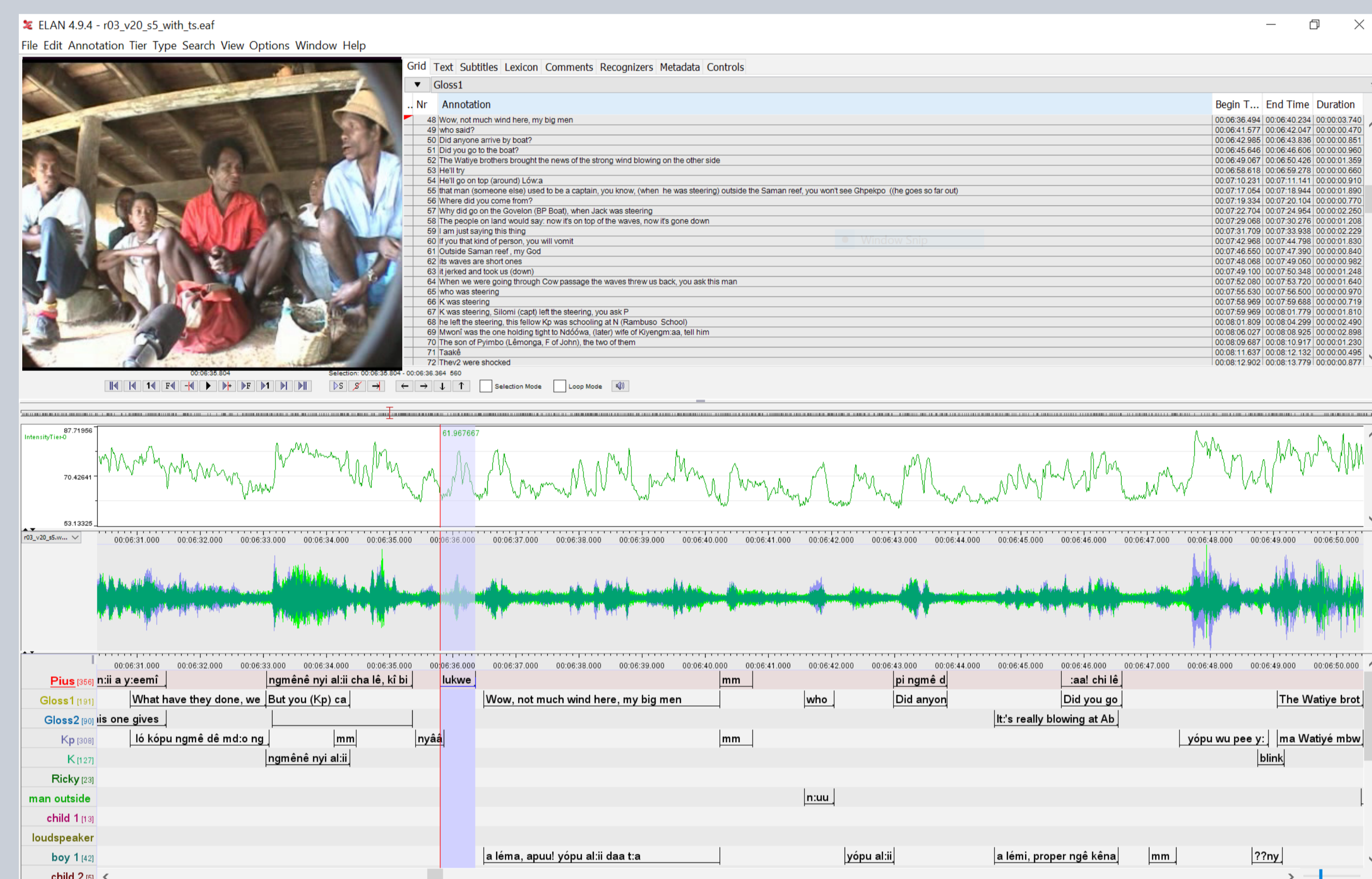


Han Slöetjes, Olaf Seibert  
han.sloetjes@mpi.nl, olaf.seibert@mpi.nl



- ELAN is a multimedia annotation tool
- freely available, sources available
- written in Java, stores data in XML
- current version 4.9.4

ELAN is a tool for multi-tier, multi-speaker, time-linked annotation of audio and video recordings



## User communities

### Language documentation

- Toolbox in/out
- FLEx in/out
- Interlinear Text out
- Praat in/out

### Sign Language

- 2 - 6 videos
- Many (>50) tiers
- External controlled vocabularies

- 1 or 2 videos, or audio only
- Segmentation
- Labelling
- Csv in/out
- Subtitles
- Simple statistics

### Behavioural research

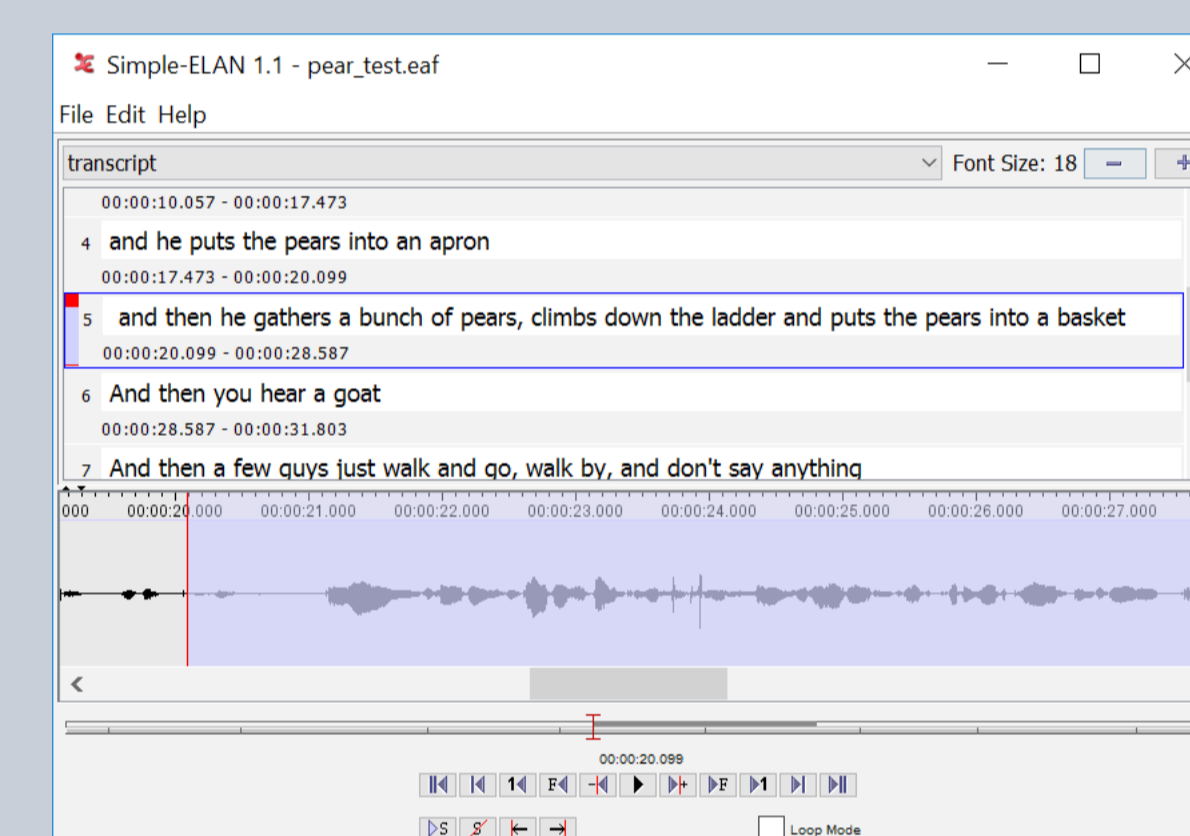
- Theme export

### Gesture/multimodality

- Annotations from overlaps, merge tiers
- Interrater reliability

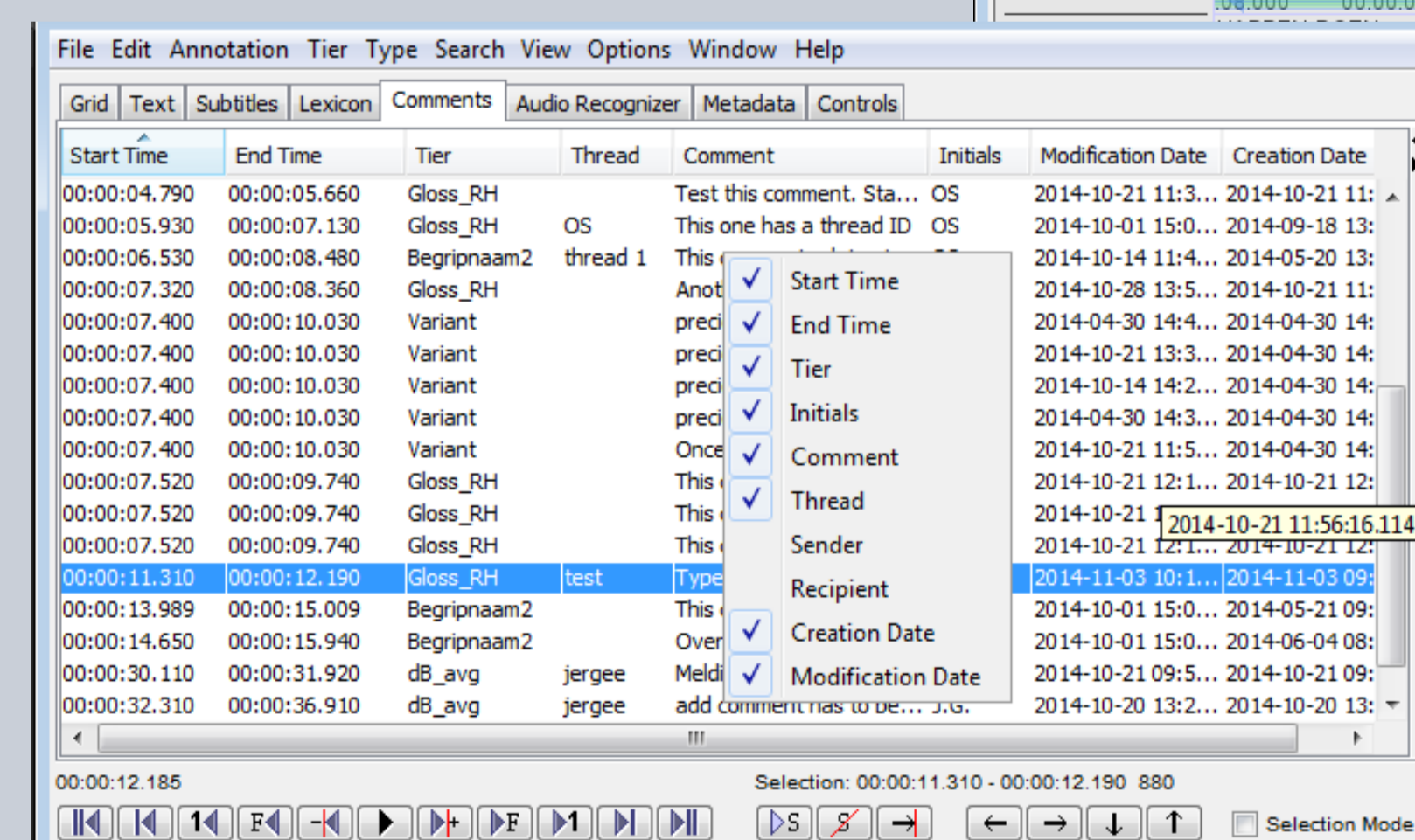
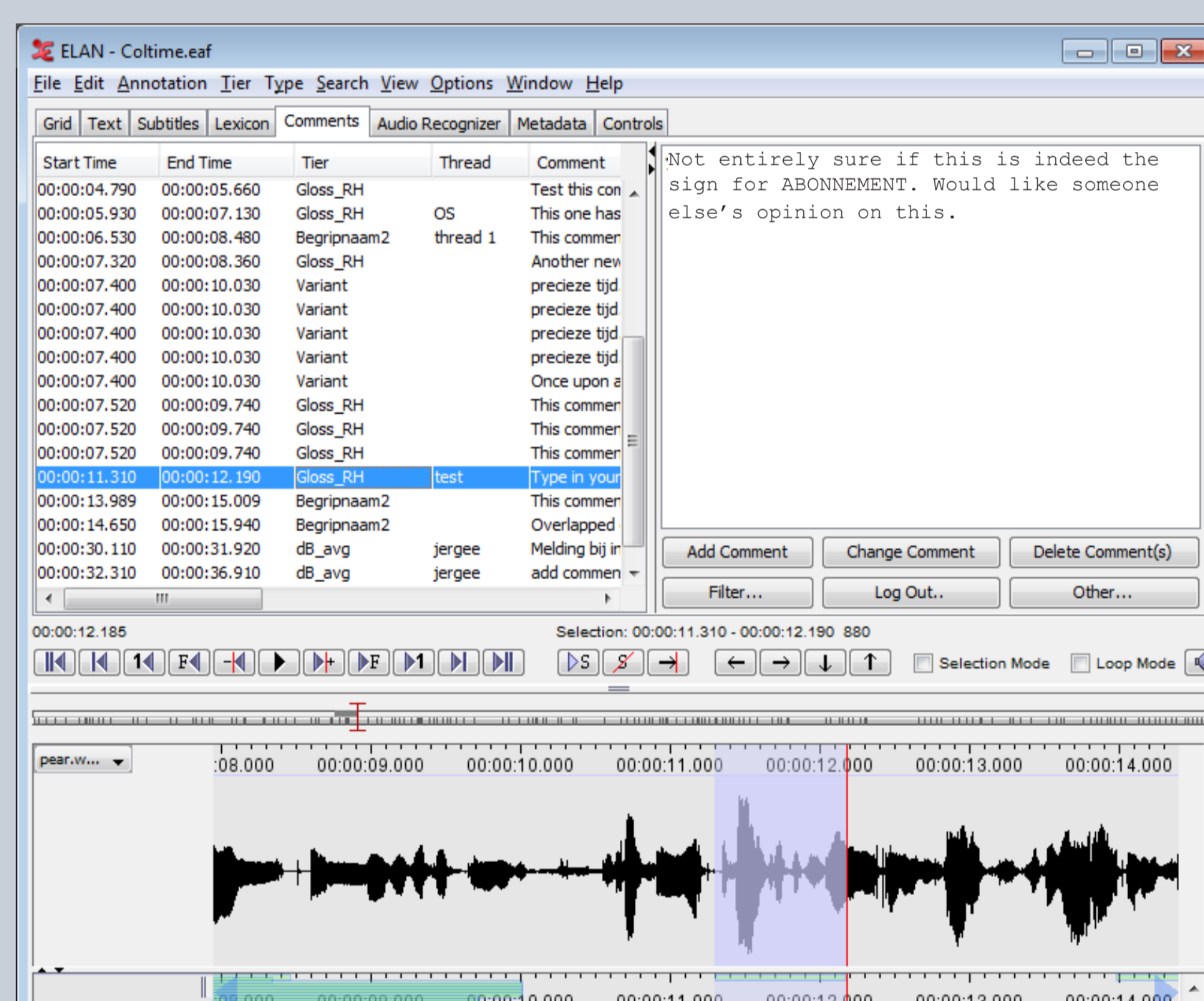
## Simple-ELAN

- Is a simplified version of ELAN
- Has limited functionality and limited complexity
- Designed for segmenting and transcribing simultaneously

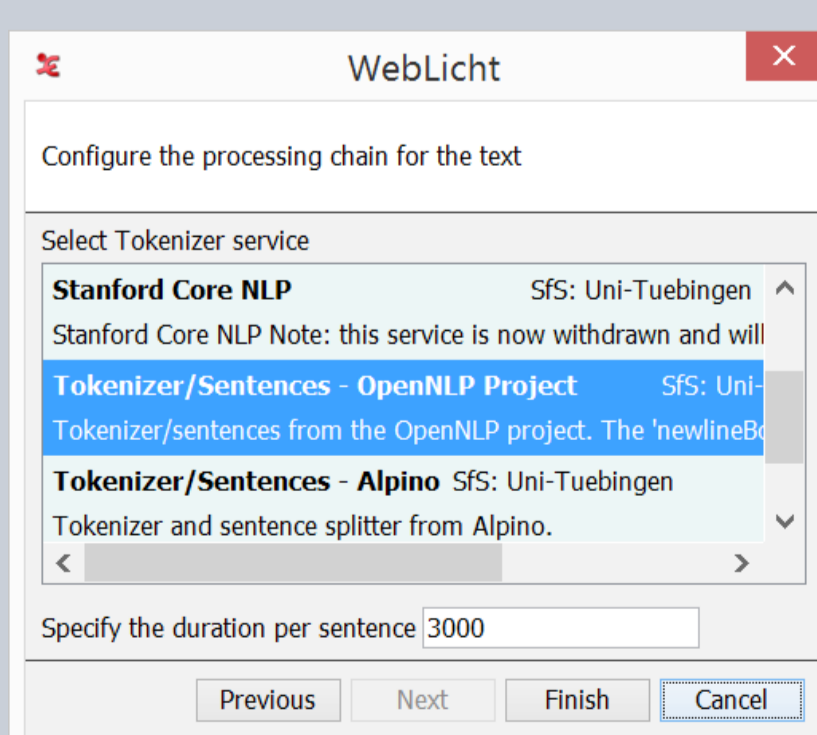


## Comments framework: sharing comments and notes with colleagues

- Comments are pieces of text linked to a segment of the media and possibly to a tier
- Can be used for notes, remarks or questions
- Share via email, a file sharing (cloud) service and/or via the back-end of the DASISH Web Annotator (DWAN)



- Filter and search comments
- Export comments
- Show/hide comments attributes



## Connect to web services

- WebLicht - select one of the available service for automatic parsing and glossing
- WebMAUS - forced alignment of text to speech signals
- Associate elements of a transcription to an ISO 639-1/3 language identifier



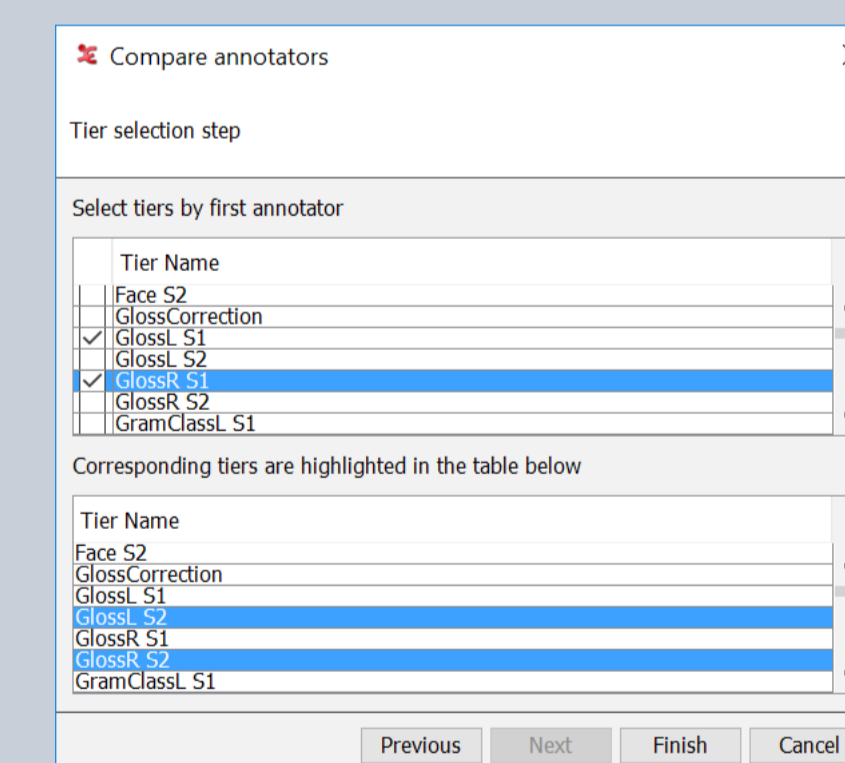
## Assessing interrater agreement

To test the validity of a coding scheme, the quality of the annotation guidelines and the training of the annotators the interrater agreement can be calculated. There seems to be no common understanding on how to assess agreement on the segmentation.

Two new algorithms have been added:

- Modified Cohen's kappa - first applies a matching algorithm to match segments created by two raters and then fills contingency tables for the kappa calculation (Holle and Rein, 2015)
- Degree of organization by applying Monte Carlo Simulations to segmentations produced by multiple raters (Lücking, 2011)

The calculation can be performed on a corpus (multiple files and multiple tiers).



## Automatic segmentation and labelling

- ELAN is a tool for manual annotation
- Manual annotation is time consuming and therefore expensive
- Attempts to implement automatic segmentation and labelling made in two projects: AVATeCH and AUVIS
- Recognizers developed so far:
  - Speech segmentation
  - Speaker diarization
  - Detection and (partial) categorization of gesture units
  - Recognize and extract key frames

