DARIAH-EU

CLARIN
Common Language Resources and
Technology Infrastructure

WP1.3 Video processing and annotation

Author:
Martin Matthiesen, CSC
Collaborators:
Mietta Lennes, UH
Juhana Salonen, JYO

Muuntamo

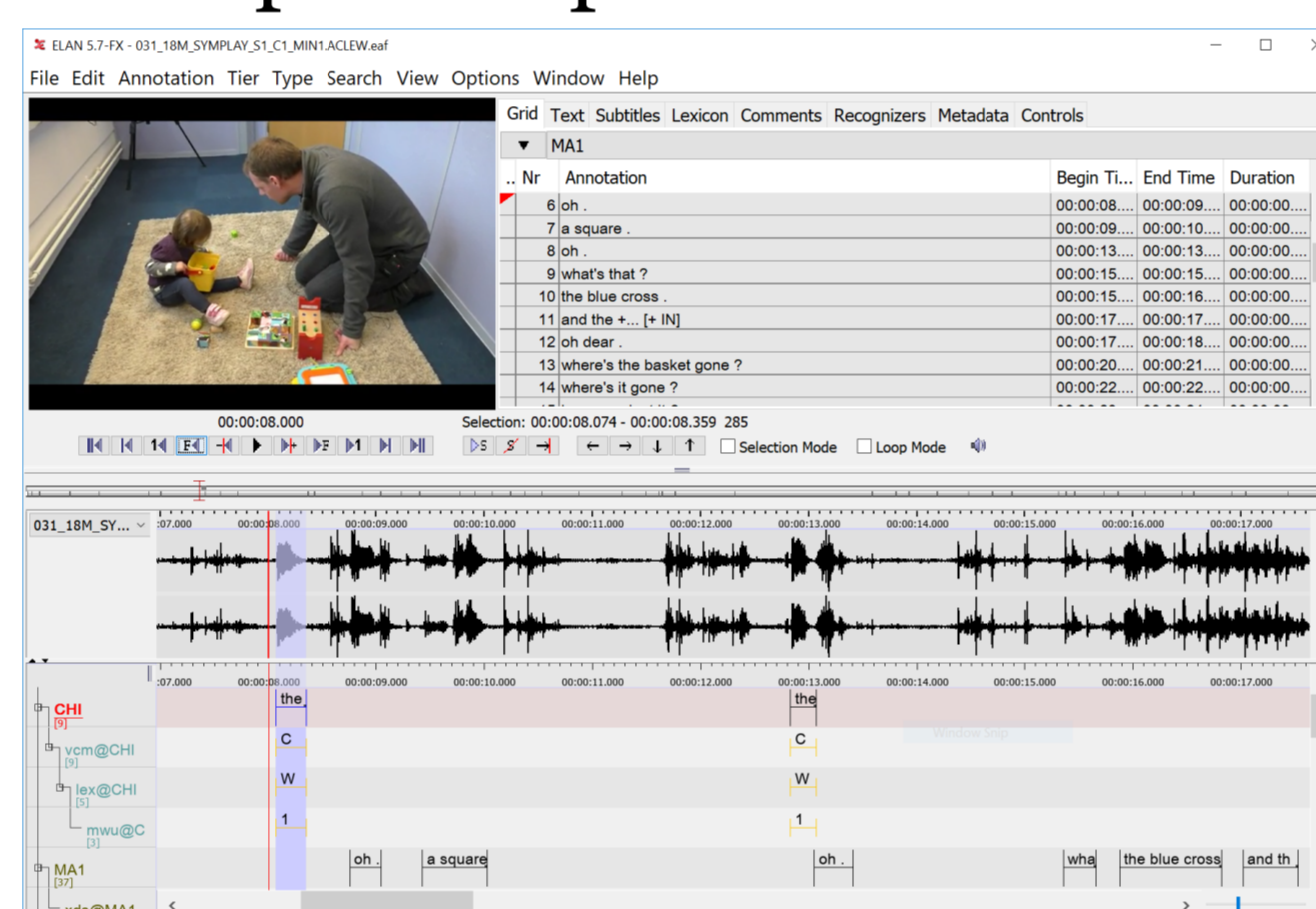## DESCRIPTION OF WORK

The foreseen **impact** is to make it easier for researchers to use, manage, annotate and share collections of video recordings as research data. We **have** corpora in Finnish Sign Language and Sessions of the Parliament of Finland, and many research groups are depositing audiovisual datasets with the LBF. In addition to automatic transcription services for audio developed in WP 1.2 and download facilities for multimedia datasets, furthering Goal 3, we **need** (D:1.3.1) to provide guidelines and access to interoperable tools for processing and annotating videos.

TIMELINE

Quarter	1	2	3	4	5	6	7	8
						D 1		

Status and Plan

Easy sharing of data requires some level of normalization. FIN-CLARIN already encourages the use of preferred formats (https://clarin.ids-mannheim.de/standards/views/view-centre.xq?id=FIN-CLARIN). Tools like ELAN and FFmpeg have been available for a long time, but their efficient large scale and secure use in HPC as well as the use in dedicated secured environments like CSC's SD-Desktop are not as well documented for users not familiar with HPC and/or secure processing environments.

We will also investigate, whether we can make existing large scale processing environments like the one used by the National Audiovisual Institute (https://kavi.fi/en/) closer to research use.

OUTCOME / Deliverable #1

- Improved documentation for ELAN/FFmpeg for use in HPC and SD-Desktop also from a data protection perspective.
- Report on possibilities of using KAVI's existing tools for research use

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

FFmpeg

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

CSC

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

DARIAH-FI

KIELIPANKKI
The Language Bank of Finland

FIN-CLARIN