ELAN
as a tool for oral history

CLARIN Oral History Workshop
Oxford

Sebastian Drude
CLARIN ERIC

18 April 2016
ELAN as a tool for oral history

1. Introducing the ELAN tool
2. Overview of basic features
3. Advanced features
4. Desiderata
5. Oral history
1. Introducing the ELAN tool

- Around 15 years of development
- Developed at the Max-Planck-Institute for Psycholinguistics in Nijmegen and by partners
- Main developer Han Sloetjes (more than 10 years)
- Used in particular in the context of:
  - Language documentation (DOBES)
  - Language acquisition studies (MPI, CHILDES)
  - Gesture studies
  - Sign language studies
1. Introducing the ELAN tool
2. Overview of basic features

• Manual annotation of audio and video recordings
• Several video streams are possible (dialogues…)
• Unlimited tiers for annotating all kind of information (“linguistic types” go beyond linguistics)
• Tiers can repeat for unlimited participants
• Hierarchy of tiers; tiers can inherit time marks
• Different input methods (support for various scripts)
• Multi-tier regular expression search
• Several in- and export formats, print
2. Overview of basic features
2. Overview of basic features
2. Overview of basic features

Tier ‘Stereotypes’

• **none**: segments can be freely set (root mother tier):
  *e.g.*, *transcription on “sentence” level*

• **symbolic association**: segments have identical start-and endpoints as in the mother tier:
  *e.g.*, *translation on “sentence” level, glosses for individual word forms or morphs*

• **included in**: segments can only be created within the boundaries of a segment on the mother tier:
  *e.g.*, *certain constituents of a sentence*
2. Overview of basic features

Tier ‘Stereotypes’

- **time subdivision**: a segment on the mother tier is completely covered by one or several adjacent segments; inner boundaries can be freely set: *e.g.*, *individual word forms*

- **symbolic subdivision**: a segment on the mother tier is completely covered by one or several adjacent segments; inner boundaries are not related to times: *e.g.*, *individual morphs within word forms*
2. Overview of basic features

Twin-tool ANNEX (online visualization tool)
3. Advanced features

Additional modes:
- Segmentation mode, transcription mode

Interaction with online services:
- Alignment of Transcriptions and Audio (MAUS)
- Automatized recognition of speech vs. silence, speaker differentiation
- Automatized recognition of moving hands & heads

ELAN Annotation Format in workflow tools
4. Desiderata

Flexibility creates a need for standards:

• Which type of information is in a tier? (tier naming)
• How is the information encoded? (conventions, abbreviations)

Independent of ELAN:

• Automatized transcription (speech recognition)
• Applying NLP-tools to (originally) oral data (named entity recognition, topic extraction…)

CLARIN
5. Oral history

• Useful for manual annotation: transcription, notes, tags
• Using advanced features, automatization (e.g. alignment of existing transcripts)
• What kinds of annotation are typically needed?
• Template for specific needs for O.H.?
• Are there technical opportunities to improve ELAN for O.H.?
• Specific metadata for O.H.?
ELAN

as a tool for oral history

CLARIN Oral History Workshop
Oxford

Sebastian Drude
CLARIN ERIC

18 April 2016