

# ELAN

## Multimedia Annotation Tool

Max-Planck-Institute for Psycholinguistics

<http://www.lat-mpi.eu/tools/elan>



Han Sloetjes ([han.sloetjes@mpi.nl](mailto:han.sloetjes@mpi.nl))

Augsburg, 30 July 2009

---

## ELAN

---

- written in Java programming language
- stores transcriptions in XML format (.eaf)
- available for Windows, Mac OS X, Linux
- sources available (GPL)
- current version 3.7.2
- Download and subscribe to mailing list:  
<http://www.lat-mpi.eu/tools/elan/download>



Release history:  
<http://www.lat-mpi.eu/tools/elan/release-notes.html>

---

# Main window

The screenshot illustrates the Elan software interface with several labeled components:

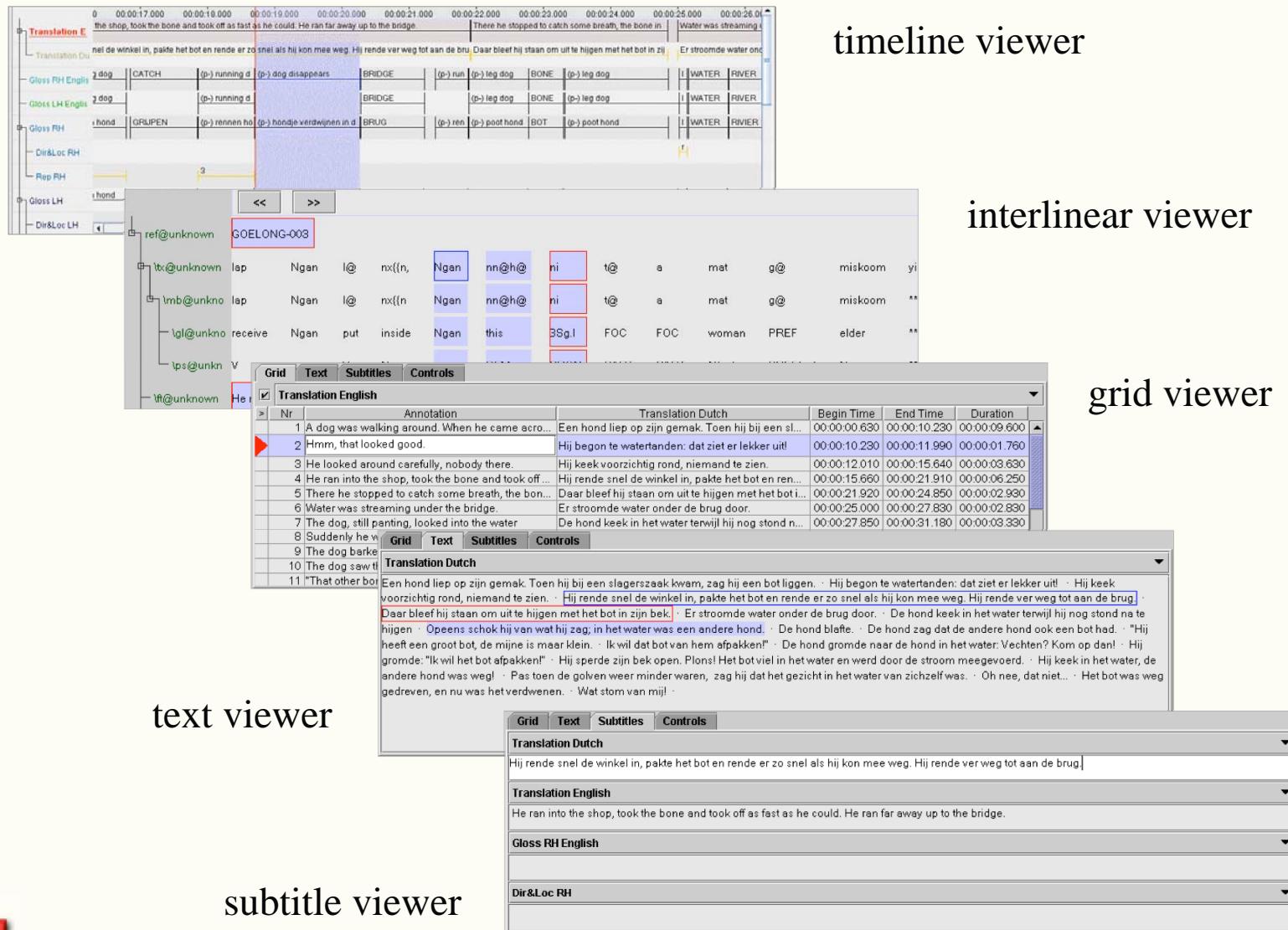
- video player**: A video frame showing three people in a hut.
- controls**: A toolbar with playback controls (rewind, fast forward, play, stop) and selection mode buttons.
- crosshair**: A red crosshair cursor pointing at the waveform viewer.
- tiers**: A list of tiers on the left: Gloss1, Gloss2, Kp, Pius, K, Ricky, man outside vie, child 1, and loudspeaker.
- viewers & controls**: A large red box covering the top right of the interface, containing:
  - Grid**, **Text**, **Subtitles**, and **Controls** tabs (with **Controls** selected).
  - A table titled "Gloss2" listing annotations with columns: Nr, Annotation, Begin Time, End Time, Duration.
  - Selection range: 00:01:01.115 - 00:01:02.182 1067.
  - Mode buttons: Selection Mode, Loop Mode.
- waveform viewer**: A timeline showing multiple audio waveforms.
- annotations**: A red box highlighting annotations in the timeline viewer, including "You have sung them, sing t", "You sing those two part", "You (P) ask him (Kp)", and "that thing, you sung those 2 parts of Myââ,".
- timeline viewer**: A timeline showing the progression of annotations over time.

A red arrow points from the "Annotations" label to the "You sing those two part" annotation in the timeline viewer.

A red arrow points from the "Tiers" label to the "child 1" tier in the list on the left.

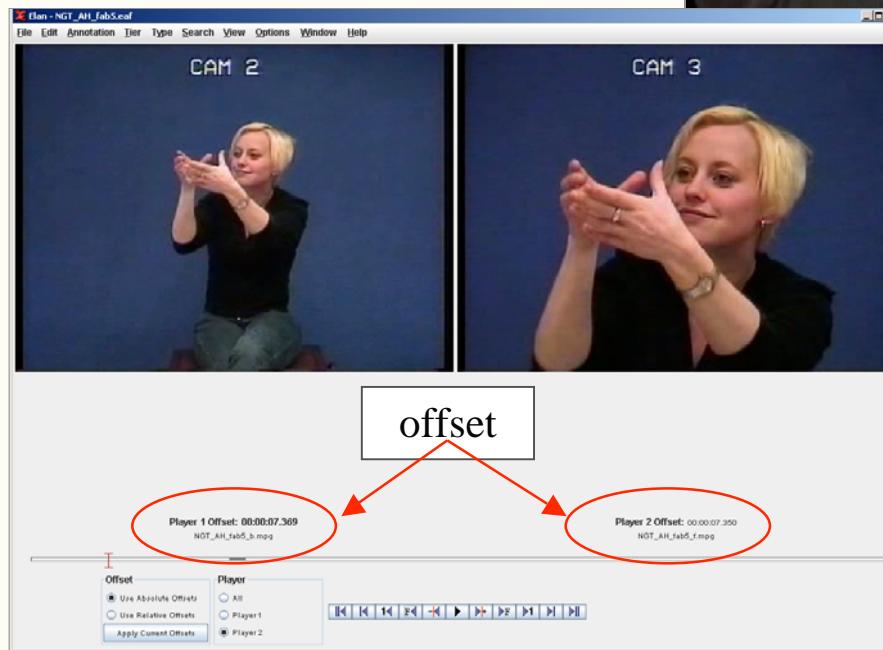
A red arrow points from the "Main window" label to the overall interface.

# Synchronized viewers



# Multiple videos

- display of up to 4 (detachable) videos



- synchronize media; determine an offset per video



---

## Media handling

---

The platform dependent part

- Windows:
  1. Windows Media Player (Direct Show, via native library)
  2. Quick Time (QT for Java, library by Apple)
  3. Java Media Framework (Sun Microsystems)
- Mac OS X:
  1. Quick Time (QT for Java, Cocoa/QTKit to Java bridge)
- Linux:
  1. Java Media Framework

Supported file types depends on the framework



---

## Key concepts

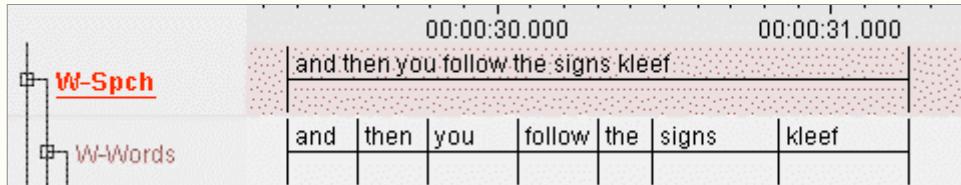
---

Tiers, linguistic types, annotations, controlled vocabularies

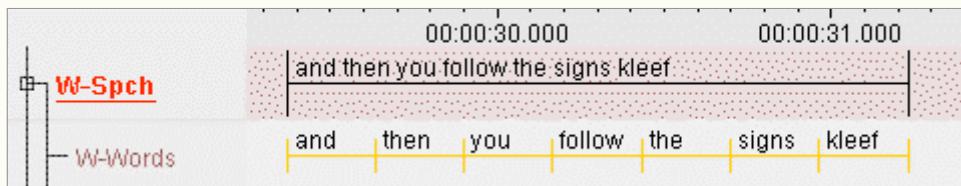
- a tier logically groups annotations that share the same constraints on structure, content and / or time alignment characteristics
- tiers can be part of a tier hierarchy, a tier can have a parent tier and dependent tiers
- an annotation document can contain any number of tiers and annotations
- a controlled vocabulary is a list of predefined annotation values on a tier



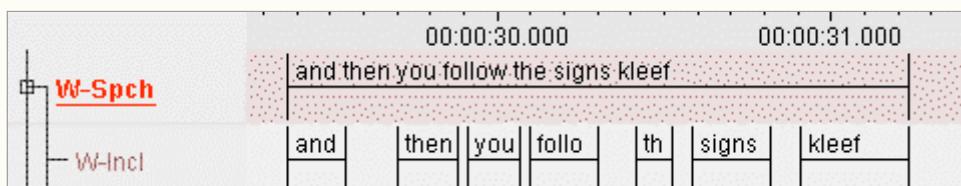
# Tiers-tier relations



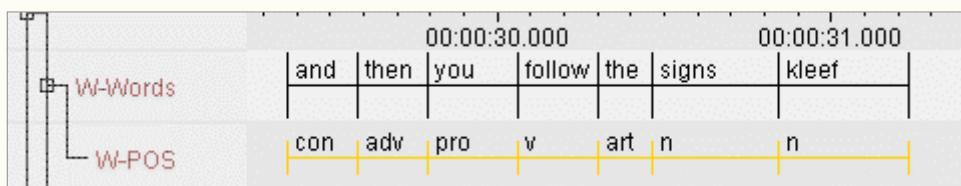
- “Time Subdivision”  
Time aligned subdivision of parent interval, no gaps



- “Symbolic Subdivision”  
Ordered, symbolic subdivision of parent interval



- “Included In”  
Time aligned subdivision of parent interval, gaps are allowed



- “Symbolic Association”  
One-to-one association with the parent annotation



# Controlled Vocabularies

- user definable lists of possible annotation values

The screenshot illustrates the process of managing a Controlled Vocabulary (CV) and applying it to a speech recording.

**CV name:** A box labeled "CV name" points to the "CV Name" field in the "Edit Controlled Vocabulary" dialog, which contains the value "phases".

**CV entries:** A box labeled "CV entries" points to the "Entries" section of the dialog, where a list of entries is shown, including "preparation", "stroke", and "hold". The entry "hold" is selected and highlighted with a blue background.

**import a CV:** A box labeled "import a CV" points to the "Import CV" button in the dialog.

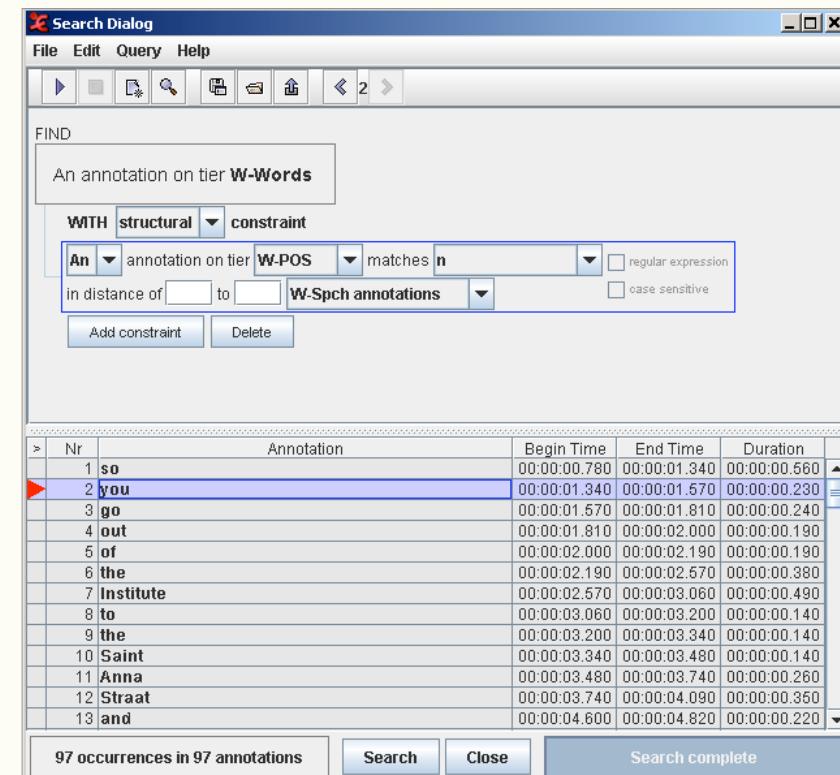
**select an entry from the list:** A box labeled "select an entry from the list" points to the "hold" entry in the list of CV entries.

To the right of the dialog, a waveform visualization shows a speech recording with several annotations overlaid. One annotation, "hold", is highlighted with a blue selection bar, corresponding to the "hold" entry in the CV list. The waveform shows the audio signal with time markers at 00:00:03.000, 00:00:04.000, and 00:00:05.000.



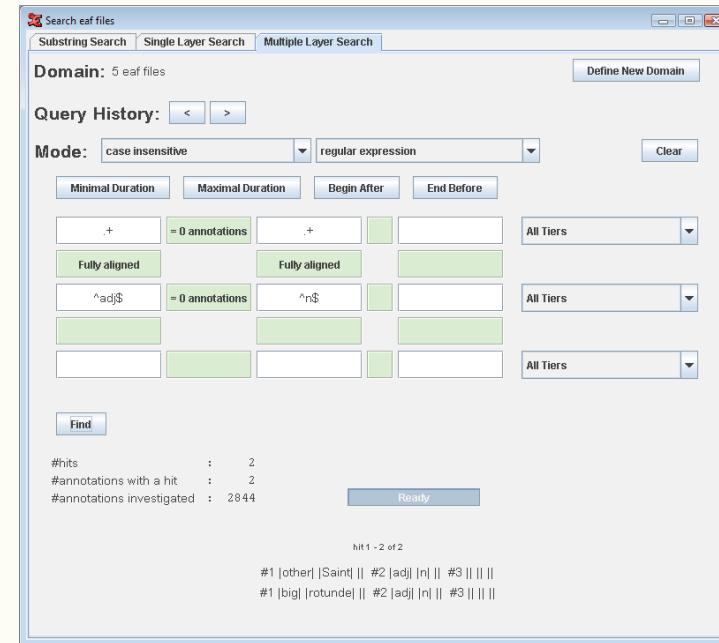
## Single file search

- create complex search queries
- combination of temporal and structural constraints
- use of regular expressions
- save and load queries
- export result to tab-delimited text
- search and replace



# Multiple file search

- add ELAN files to a 'basket'
- simple and structured search: several matching types, in selected set of tiers
- concordance and frequency view
- export results to tab-delimited text file
- open a file by a click on a result



Annotations								
Search going				Define search domain	Search	Export	Search complete	
Nr	File	Tier	Before	Annotation	After	Begin Time	End Time	Duration
1	elan-demo	W-RGMe		Going out of the Institute	Going out of MPI to...	00:00:00.780	00:00:04...	00:00:03...
2	elan-demo	W-RGMe	Going out of the...	Going out of MPI to Saint	Going along St. An...	00:00:04.600	00:00:10...	00:00:05...
3	elan-demo	W-RGMe	Going out of M...	Going along St. Anna Street	Going along St. An...	00:00:31.320	00:00:31...	00:00:00...
4	elan-demo	W-RGMe	Going along St...	Going along St. Anna Street	A roundabout	00:00:32.379	00:00:34...	00:00:01...
5	elan-demo	W-RGMe	A roundabout	Going to the direction of	Going along Oran...	00:00:01.570	00:00:01...	00:00:00...
6	elan-demo	W-RGMe	Going to the dire...	Going along Orange Singel	Going to the directi...	00:00:05.370	00:00:05...	00:00:00...
7	elan-demo	W-RGMe	Going along Or...	Going to the direction of	Going down	00:00:31.320	00:00:31...	00:00:00...
8	elan-demo	W-RGMe	Going to the dir...	Going down	a roundabout	00:00:32.920	00:00:33...	00:00:00...
9	elan-demo	W-RGMe		Going down	Going down	00:00:00.630	00:00:10...	00:00:09...
10	elan-demo	W-RGMe	Going down	Going down	a roundabout	00:00:27.850	00:00:31...	00:00:03...
11	elan-demo	W-RGMe	a roundabout	Going down	Going down	00:00:31.370	00:00:33...	00:00:02...
12	elan-demo	W-RGMe	Going down	Going down	Going to Kleve	00:00:33.820	00:00:35...	00:00:01...
13	elan-demo	W-RGMe	Going down	Going to Kleve	Going down	00:00:35.280	00:00:38...	00:00:03...
14	rabbitENG01_test0	speech		(going) hundred and eighty	backwards and for...	00:00:43.490	00:00:48...	00:00:05...
15	rabbitENG01_test0	speech	turning around i...	it's going	but meandering fr...	00:00:58.650	00:01:01...	00:00:02...
16	rabbitENG01_test0	speech	but meanderin...	(going) forward to it	for a few steps	00:01:01.570	00:01:11...	00:00:09...
17	rabbitENG01_test0	speech	turning round b...	(going) in a kind of loop	It's coming down i...	00:00:00.640	00:00:01...	00:00:01...
18	rabbitENG01_test0	speech	but this time it's...	(going) round three	as it's moving roun...	00:00:01.920	00:00:04...	00:00:02...
19	rabbitENG01_test0	speech		(going) hundred and eighty	backwards and for...	00:00:10.430	00:00:11...	00:00:00...
20	rabbitENG01_test0	speech	turning around i...	it's going	but meandering fr...	00:00:15.670	00:00:16...	00:00:01...
21	rabbitENG01_test0	speech	but meanderin...	(going) forward to it	for a few steps	00:00:17.860	00:00:18...	00:00:00...
22	rabbitENG01_test0	speech	turning round b...	(going) in a kind of loop	It's coming down i...	00:00:18.710	00:00:20...	00:00:01...
23	rabbitENG01_test0	speech	but this time it's...	(going) round three	as it's moving roun...	00:00:21.890	00:00:22...	00:00:00...



find-and-replace in multiple files

# Statistics

- basic statistics for the annotations of a selected tier or for all tiers in the transcription
- save as tab-delimited text file

The image displays two windows of the 'Annotation Statistics' software. Both windows have a title bar 'Annotation Statistics' and a main header 'Statistics'.  
**Left Window (Annotations Tab):**  
The tab bar at the top has 'Annotations' (selected) and 'Tiers'. Below it is a section titled 'Tiers' which contains a table with one row: 'Gloss RH'. Underneath is a section titled 'Statistics Variables' containing a table with columns: Annotation, Occurrences, Frequency, and Average. The table lists various annotations with their counts and average frequencies.  

Annotation	Occurrences	Frequency	Average
(g-) "kom maar...	3	0.0372023809...	0.3833
(g-) pu	1	0.0124007936...	0.52
(p-) drijven	2	0.0248015873...	1.545
(p-) groot bot	1	0.0124007936...	1.51
(p-) hondje ver...	1	0.0124007936...	1.56
(p-) klaww hond	3	0.0372023809...	0.9833
(p-) klein bot	1	0.0124007936...	0.65
(p-) lopen hond	1	0.0124007936...	2.66
(p-) poot hond	6	0.0744047619...	1.1816

  
**Right Window (Tiers Tab):**  
The tab bar at the top has 'Annotations' and 'Tiers' (selected). Below it is a section titled 'Statistics Variables' containing a table with columns: Tier, Number of ..., Minimal Du..., Maximal Du..., Average D..., Median Dur..., Total Anno..., and Latency. The table lists various tiers with their respective statistics.  

Tier	Number of ...	Minimal Du...	Maximal Du...	Average D...	Median Dur...	Total Anno...	Latency
Gloss1	191	0.38	4.065	1.504021	1.332	287.268	14.74
Gloss2	90	0.28	3.15	1.2127	1.107	109.143	31.708
Kp	308	0.145	9.2	0.956321	0.722	294.547	4.95
Pius	356	0.25	9.81	1.375829	1.189	489.795	3.1
K	127	0.28	5.525	1.008803	0.83	128.118	1.38
Ricky	23	0.31	2.82	0.939435	0.605	21.607	257.721
man outsid...	103	0.27	2.665	0.872913	0.629	89.91	31.158
child 1	13	0.34	2.055	1.128692	0.99	14.673	15.95
loudspeaker	3	3.069	4.308	3.812333	4.06	11.437	8.353
boy 1	42	0.38	3.74	1.238738	1.205	52.027	32.988
child 2	5	0.4	2.06	1.08	0.52	5.4	32.988
boy 2	2	0.57	1.88	1.225	1.225	2.45	135.166
boy 3	9	0.52	3.74	1.315222	0.93	11.837	178.301



# LAT Tools

- tools developed at the MPI for Psycholinguistics



**Language Archiving Technology**

## Tools



Addit a tool to 'plug' notes to elements of the MPI archive.



AMS a tool to grant and deny the access to a (part of a) corpus.



Annex the Annotation Exploration tool in the MPI web-based framework for archive exploration and enrichment



Elan Multimedia Annotator



Imdi create, search through and modify metadata for linguistic resources



Lamus Language Archive Management and Upload System



Lexus a web-based lexicon tool



Sympathy a tool for manual syntactical annotation



# LAT Tools

- ANNEX/TROVA; web-based search and exploration of archive resources

The screenshot shows the MPI - ANNEX Interface. At the top, there's a browser header with the URL [http://corpus1.mpi.nl/ds/annex/protected/no\\_auth.jsp?sessionid=EC5DC3B85129432C2C2FE19EACF20882&nodeid=MPI303305%23&time=82331&du](http://corpus1.mpi.nl/ds/annex/protected/no_auth.jsp?sessionid=EC5DC3B85129432C2C2FE19EACF20882&nodeid=MPI303305%23&time=82331&du). Below the header, the interface has several sections:

- ANNEX manual ? settings**: A navigation bar.
- Text**, **Grid**, **Subtitle**, **Timeline**: Selection buttons on the left.
- Video display**: A video player showing a group of people in a traditional setting.
- Media information**: Details about the resource: MISE-1.eaf, Media file: MISE-1.mp4, Elapsed time: 00:01:21:331.
- Mini Data Frame**: A text area displaying a sentence in a language like Ilocano with English glosses: "Alukwem lova, lova ama anigada Gunter kaitala bani. Bogeokwa. Ava laweya taninua. Akau ale'i eta'ina nata yena naveaka iseki wala. Aweya aweya itotina, boreokwa. Ababivu wala, ale'i abibia ameya omatagu ikalikuvu ila. Natalaga abiu nayu e deli taninua ake'kita ama o valu akabwasi. A nata a musa natabwabogwa, kudoki, ameya itokulati wala kwena. Manasita na'uvanaku. Lova lova labia. E yokwamiga? Sita ka kweta sotoli lova kuvagis? Bogwa latamesi beya bogwa kulalasi. Kula bakayopolusiga A bwada gala bili Tavagisi pola. Meyaga pwepu abili bogwaga o kuvakanotsasi. A bibili tombaiku - kwalaisiga bogwa epaisewa. - ?? Ke gugwadi beya labilibili e ekikauvala gala."
- Timeline**: A grid showing text segments across different tiers (mb1, gl1, ft1, ref2, mb2, gl2, ft2). The segments are color-coded and some have yellow highlights. The text in the grid includes:
  - mb1: tu a-le'i e-ta'ina a-weya a-weya i-totina bogwa e-vokuva a-babi-vau wala a-le'i a-bi-bia a-meya omata-gu i-kalikuvu i-la na-tala-ga a-biu na-yu e
  - gl1: hed 1.-go.to 1. P; 1.-hit 1.-hit 3.-break already 3.-be.finished 3.-stick.bait.on-again only 1.-throw 1.-Redup-pull 1.-bring in.front.of-me 3.-get.away 3.-
  - ft1: . I went to catch s; I shoot the line - I shoot the line - it breaks (the hook), it is over. I just stick a new bait on it again, I shoot the line, I pull and pull, I bring it in f
  - ref2:
  - mb2:
  - gl2:
  - ft2:
- Buttons and controls: **Play selection**, **Clear selection**, navigation arrows (<>, <>>, <><>), and checkboxes for **Play screen by screen** and **Play continually**.
- Font settings: **Tier:** ref1, **Font size:** 14.



# EAF Format and Interoperability

## EAF file format

- description of EAF format and XML schema

## Interoperability

- of linguistic concepts: ISO Data Category Registry
- by means of a per tool/format import and export mechanism
- by means of an interchange or pivot format



## EAF format

The EAF document format is defined in an XML Schema, current version is 2.6.

The document format, the naming of the elements and the referential structures are fairly explicit.

At MPI separation of metadata and annotation data

Apart from what is/can be defined in the schema, ELAN puts additional constraints on the usage of elements and references.

Some basic constraints:

Annotations on any tier cannot overlap

Time-alignable annotations are identified by references to two distinct time slots

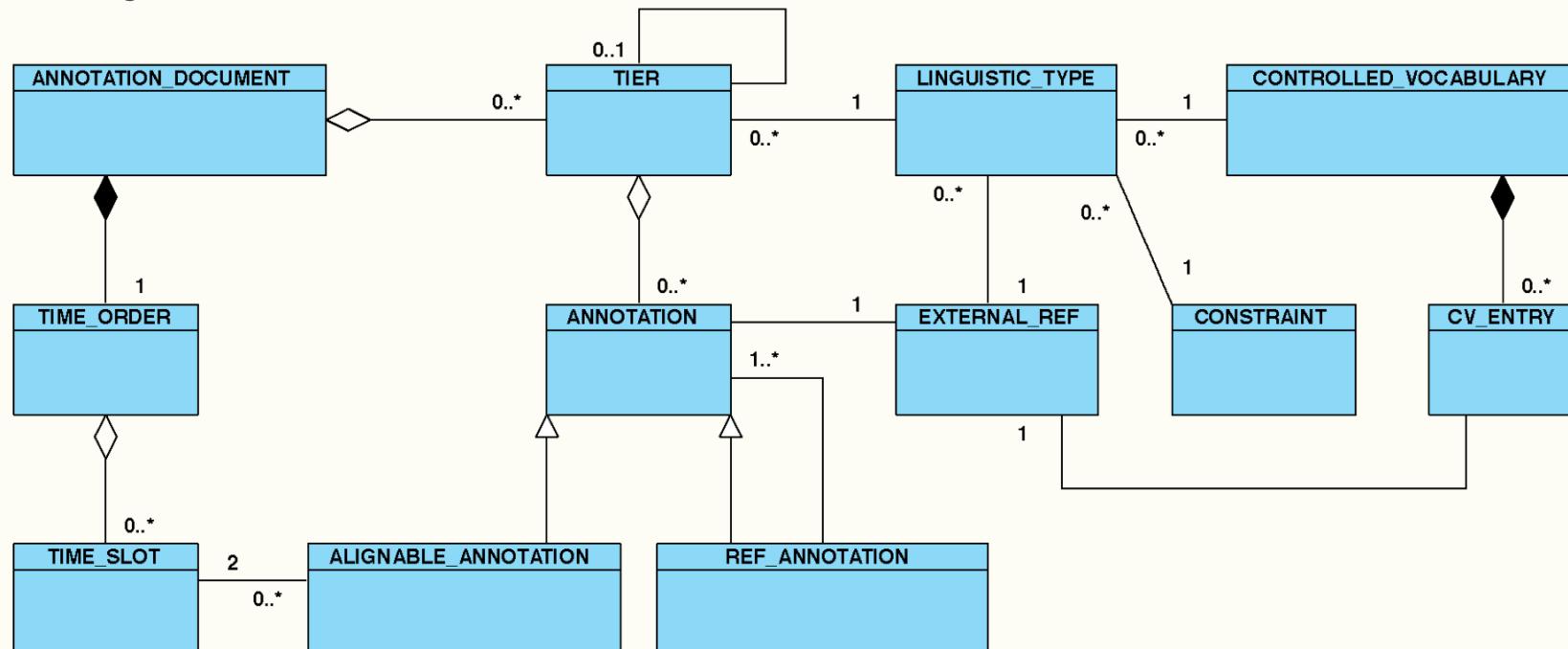
Neither ELAN nor EAF supports references to fragments of other annotations



# Class diagram

(Simplified) UML diagram of EAF documents and ELAN transcription objects.

Annotations are contained in Tier objects. A tier can be associated with a parent tier. Alignable and reference annotation are distinct objects.



---

## EAF format

---

Elements: **TIME\_ORDER** and **TIME\_SLOT**

Explicit naming: time slot instead of e.g. anchor

Duplicates of time slots with the same time value

Empty time slots allowed (no time value)

```
<TIME_ORDER>
  <TIME_SLOT TIME_SLOT_ID="ts1" TIME_VALUE="1533"/>
  <TIME_SLOT TIME_SLOT_ID="ts2" TIME_VALUE="3495"/>
  <TIME_SLOT TIME_SLOT_ID="ts3" TIME_VALUE="3576"/>
  <TIME_SLOT TIME_SLOT_ID="ts4" TIME_VALUE="5166"/>
  <TIME_SLOT TIME_SLOT_ID="ts5" TIME_VALUE="5166"/>
  <TIME_SLOT TIME_SLOT_ID="ts6" TIME_VALUE="6444"/>
  <TIME_SLOT TIME_SLOT_ID="ts7" TIME_VALUE="6444"/>
  <TIME_SLOT TIME_SLOT_ID="ts8"/>
  <TIME_SLOT TIME_SLOT_ID="ts9"/>
  <TIME_SLOT TIME_SLOT_ID="ts10" TIME_VALUE="8940"/>
  <TIME_SLOT TIME_SLOT_ID="ts11" TIME_VALUE="8940"/>
  <TIME_SLOT TIME_SLOT_ID="ts12" TIME_VALUE="9381"/>
</TIME_ORDER>
```



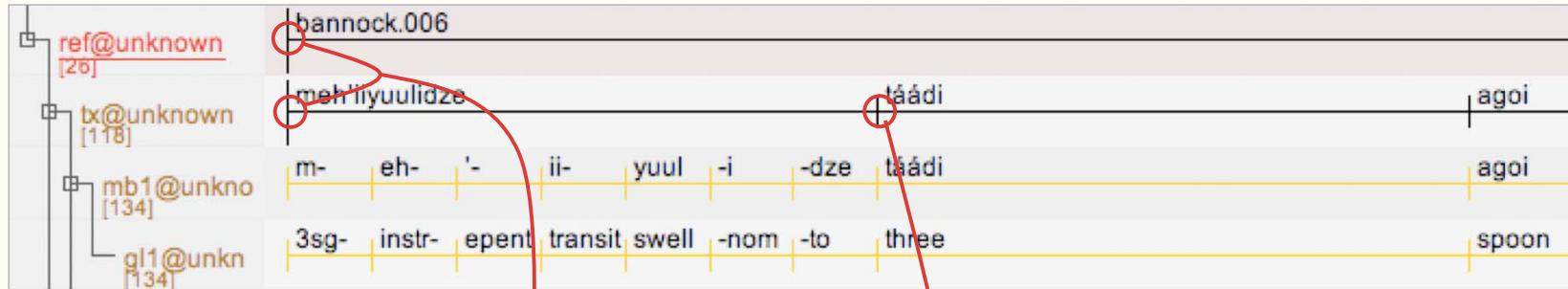
# EAF format

Usage of Time Slots:

Shared references: chaining of annotations on the same tier or between tiers

Slots without time value for unaligned annotations on alignable tiers

Duplicates of slots with the same time value



Unaligned annotation, referring to a time slot without time value (only on Time Subdivision tiers)

Two annotations on unrelated tiers with the same begin time, 2 time slots with the same value



---

# EAF format

---

## Elements: TIER and ANNOTATION

Tier name is also the id of the tier

A tier can refer to a parent tier: this puts constraints on its type:  
(top level tier -> linguistic type=Stereotype "None", other tier ->  
linguistic type=Stereotype not "None")

Alignable and Reference annotations are embedded in Annotation  
elements, the annotation value is a separate element

```
<TIER TIER_ID="Clause Transcript" LINGUISTIC_TYPE_REF="Motion" DEFAULT_LOCALE="en">
  <ANNOTATION>
    <ALIGNABLE_ANNOTATION ANNOTATION_ID="a2" TIME_SLOT_REF1="ts2" TIME_SLOT_REF2="ts5">
      <ANNOTATION_VALUE>so it starts out with a rooster crows</ANNOTATION_VALUE>
    </ALIGNABLE_ANNOTATION>
  </ANNOTATION>
</TIER>
<TIER TIER_ID="Motion" LINGUISTIC_TYPE_REF="Sp Comm" PARENT_REF="Clause Transcript">
  <ANNOTATION>
    <REF_ANNOTATION ANNOTATION_ID="a18" ANNOTATION_REF="a2">
      <ANNOTATION_VALUE>Non-motion</ANNOTATION_VALUE>
    </REF_ANNOTATION>
  </ANNOTATION>
</TIER>
```

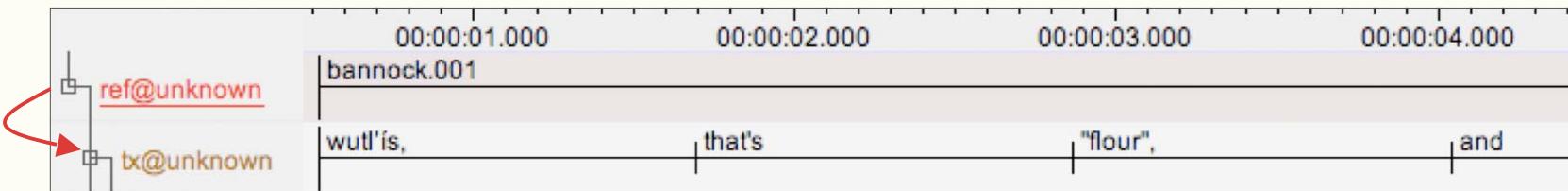


## EAF format

Alignable annotations: 2 time slots are required.

Annotations on dependent tiers are always within the interval of the parent annotation.

On **Time Subdivision tier**: first annotation has the same time slot as the parent annotation, the second time slot is the first slot of the next annotation etc. (no gaps). The last annotation of the chain shares the end slot with the parent annotation



```
<TIER TIER_ID="ref@unknown" PARTICIPANT="unknown" LINGUISTIC_TYPE_REF="ref">
  <ANNOTATION>
    <ALIGNABLE_ANNOTATION ANNOTATION_ID="a1" TIME_SLOT_REF1="ts1" TIME_SLOT_REF2="ts6">
      <ANNOTATION_VALUE>bannock.001</ANNOTATION_VALUE>
    </ALIGNABLE_ANNOTATION>
  </ANNOTATION>
  ...
<TIER TIER_ID="tx@unknown" PARTICIPANT="unknown" LINGUISTIC_TYPE_REF="tx"
      PARENT_REF="ref@unknown">
  <ANNOTATION>
    <ALIGNABLE_ANNOTATION ANNOTATION_ID="a27" TIME_SLOT_REF1="ts1" TIME_SLOT_REF2="ts2">
      <ANNOTATION_VALUE>wutl's,</ANNOTATION_VALUE>
    </ALIGNABLE_ANNOTATION>
  </ANNOTATION>

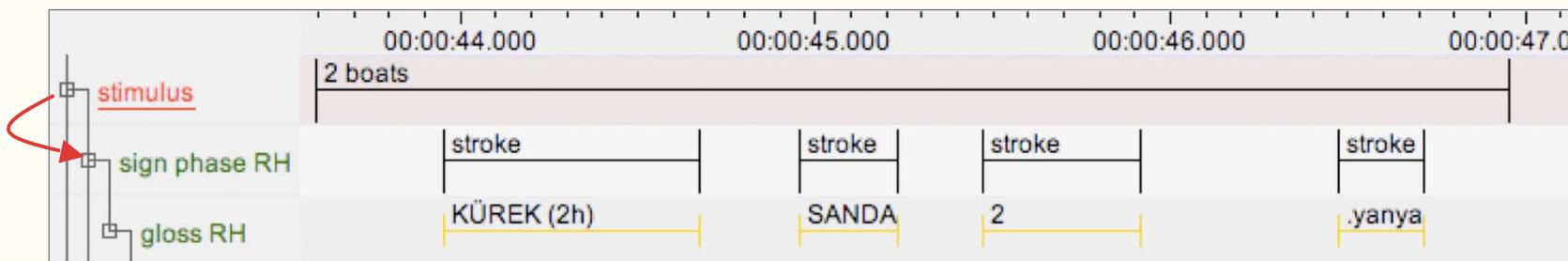
```



# EAF format

On top-level tiers time slots are not shared.

On **Included In** dependent tiers: gaps are allowed, each annotation refers to a begin time slot and end time slot that is not shared with any other annotation.



```
<TIER DEFAULT_LOCALE="nl" LINGUISTIC_TYPE_REF="default-lt" TIER_ID="stimulus">
  <ANNOTATION>
    <ALIGNABLE_ANNOTATION ANNOTATION_ID="a61" TIME_SLOT_REF1="ts33" TIME_SLOT_REF2="ts46">
      <ANNOTATION_VALUE>2 boats</ANNOTATION_VALUE>
    </ALIGNABLE_ANNOTATION>
  </ANNOTATION>
  ...
<TIER DEFAULT_LOCALE="en" LINGUISTIC_TYPE_REF="sign phase included in" PARENT_REF="stimulus"
      TIER_ID="sign phase RH">
  <ANNOTATION>
    <ALIGNABLE_ANNOTATION ANNOTATION_ID="a62" TIME_SLOT_REF1="ts34" TIME_SLOT_REF2="ts36">
      <ANNOTATION_VALUE>stroke</ANNOTATION_VALUE>
    </ALIGNABLE_ANNOTATION>
  </ANNOTATION>
  ...

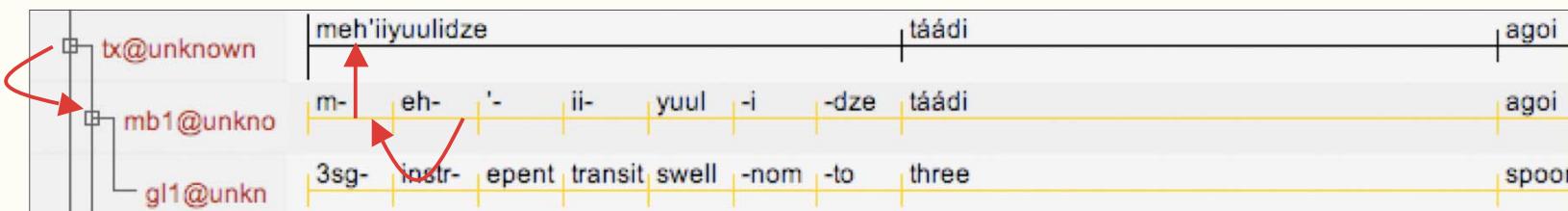
```



# EAF format

Reference annotations refer to a parent annotation.

In addition, on symbolic subdivision tiers annotations also refer to the previous annotation in the chain of depending annotations, except for the first annotation in the chain).



```
<ANNOTATION>
  <REF_ANNOTATION ANNOTATION_ID="a161" ANNOTATION_REF="a60">
    <ANNOTATION_VALUE>m-</ANNOTATION_VALUE>
  </REF_ANNOTATION>
</ANNOTATION>
<ANNOTATION>
  <REF_ANNOTATION ANNOTATION_ID="a162" ANNOTATION_REF="a60" PREVIOUS_ANNOTATION="a161">
    <ANNOTATION_VALUE>eh-</ANNOTATION_VALUE>
  </REF_ANNOTATION>
</ANNOTATION>
<ANNOTATION>
  <REF_ANNOTATION ANNOTATION_ID="a163" ANNOTATION_REF="a60" PREVIOUS_ANNOTATION="a162">
    <ANNOTATION_VALUE>'</ANNOTATION_VALUE>
  </REF_ANNOTATION>
</ANNOTATION>
```



---

# EAF format

---

## Elements: **LINGUISTIC TYPE** and **CONSTRAINT**

The name of a linguistic type is its id

The TIME\_ALIGNABLE attribute should be consistent with the CONSTRAINT reference

The CONSTRAINTS attribute is optional (unfortunately)

There are 4 predefined “stereotypes” (the 5th “multiple\_refs” not yet implemented)

```
<LINGUISTIC_TYPE LINGUISTIC_TYPE_ID="ref" TIME_ALIGNABLE="true"/>
<LINGUISTIC_TYPE LINGUISTIC_TYPE_ID="id" TIME_ALIGNABLE="false" CONSTRAINTS="Symbolic_Association"/>
<LINGUISTIC_TYPE LINGUISTIC_TYPE_ID="tx" TIME_ALIGNABLE="true" CONSTRAINTS="Time_Subdivision"/>
<LINGUISTIC_TYPE LINGUISTIC_TYPE_ID="mb1" TIME_ALIGNABLE="false"
    CONSTRAINTS="Symbolic_Subdivision"/>

<CONSTRAINT STEREOTYPE="Time_Subdivision" DESCRIPTION="Time subdivision of parent annotation's time
interval, no time gaps allowed within this interval"/>
<CONSTRAINT STEREOTYPE="Symbolic_Subdivision" DESCRIPTION="Symbolic subdivision of a parent annotation.
Annotations referring to the same parent are ordered"/>
<CONSTRAINT STEREOTYPE="Symbolic_Association" DESCRIPTION="1-1 association with a parent annotation"/>
<CONSTRAINT STEREOTYPE="Included_In" DESCRIPTION="Time alignable annotations within the parent annotation's
time interval, gaps are allowed"/>
```



## EAF format

Summary:

Top-level tier: linguistic type without a constraint attribute, annotations are Alignable Annotations, annotations cannot overlap, gaps are allowed (no sharing of time slots)

Time Subdivision tier: linguistic type with constraint of stereotype Time\_Subdivision, annotations are Alignable Annotations, no gaps allowed, chaining of annotations by sharing of time slots

Included In tier: linguistic type with constraint of stereotype Included\_In, annotations are Alignable Annotations, gaps are allowed, no sharing of time slots

Symbolic Subdivision tier: linguistic type with constraint of stereotype Symbolic\_Subdivision, annotations are Ref Annotations, reference to a parent annotation required, chaining by reference to previous annotation

Symbolic Association tier: linguistic type with constraint of stereotype Symbolic\_Association, annotations are Ref Annotations, reference to a parent annotation is required, maximal one child annotation per parent



# EAF format

Elements: **CONTROLLED VOCABULARY** and **EXTERNAL REF**

Currently the only allowed type for the external reference element is "iso12620" (ISO DCR).

Annotations, Linguistic Types and Controlled Vocabulary entries can be associated with a data category.

```
<LINGUISTIC_TYPE LINGUISTIC_TYPE_ID="Character vpt" TIME_ALIGNABLE="false"  
CONSTRAINTS="Symbolic_Association" CONTROLLED_VOCABULARY_REF="Gs Hand Shape"/>  
  
<CONTROLLED_VOCABULARY CV_ID="Gs Hand Shape" DESCRIPTION="">  
  <CV_ENTRY DESCRIPTION="">clutched</CV_ENTRY>  
  <CV_ENTRY DESCRIPTION="">open</CV_ENTRY>  
  <CV_ENTRY DESCRIPTION="">unclear</CV_ENTRY>  
</CONTROLLED_VOCABULARY>  
  
<CONTROLLED_VOCABULARY CV_ID="POS" DESCRIPTION="Part of Speech">  
  <CV_ENTRY DESCRIPTION="noun" EXT_REF="er1">n</CV_ENTRY>  
  ...  
  
<EXTERNAL_REF EXT_REF_ID="er1" TYPE="iso12620" VALUE="1333"/>
```

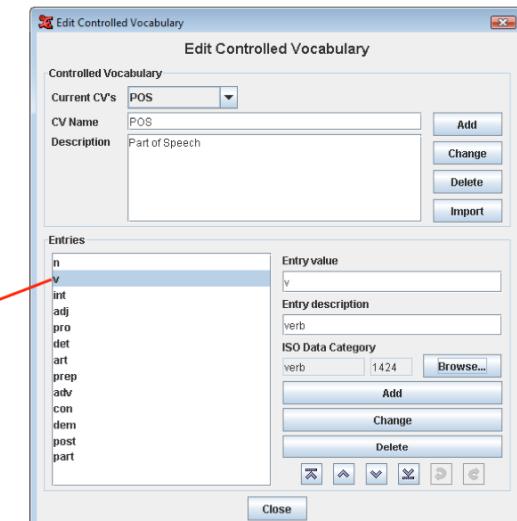
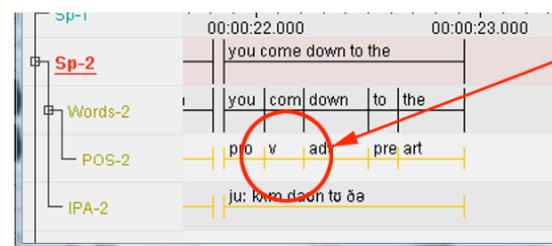


## Interaction with ISO DCR

The ISO Data Category Registry defines linguistic concepts; their definitions and descriptions. It provides standards for Language Resource management, creation and coding.  
<http://www.isocat.org/>

In ELAN on the level of

- annotations
- linguistic types
- entries in Controlled Vocabularies



## Import / export

format	import	export
Shoebox/Toolbox (.txt)	✓	✓
Transcriber (.trs)	✓	
Chat (.cha)	✓	✓
Praat (.TextGrid)	✓	✓
CSV/tab-delimited text	✓	✓
FLEx	✓	
Interlinear text		✓
HTML		✓
Wordlist		✓
Subtitles (SMIL, QT, SubRip)		✓



## AG interchange format

Initiative carried out during two ISGS conference workshops, 7 tools involved

Based on the Annotation Graph xml format (ag-1.1.dtd)

EAF implementation based on the AGLib toolkit,  
unfinished...

Tier information stored in Metadata elements...

Retaining the basic annotation information is relatively straightforward, main problem for full roundtrip is loss of tier structure information, speaker assignment etc



## GrAF interchange format

XML implementation of the ISO LAF (Linguistic Annotation Framework) initiative. In development, current schema version 1.0.6b?

No import/export from ELAN yet.

Export probably not problematic, can be done in several ways?

Import unsure...

Retaining tier structure information likely equally problematic as with AG.



# ELAN

<http://www.lat-mpi.eu/tools/elan>

E. Auer, H. Brugman, G. Gulrajani, A. Klassmann, A. Koenig, M. Kramer,  
M. Pippel, A. Russel, H. Sloetjes, H. Spenke



Han Sloetjes

Augsburg, 30 July 2009

## Import / export

format	import	export
Shoebox/Toolbox (.txt)	✓	✓
Transcriber (.trs)	✓	
Chat (.cha)	✓	✓
Praat (.TextGrid)	✓	✓
CSV/tab-delimited text	✓	✓
FLEX	✓	
Interlinear text		✓
HTML		✓
Wordlist		✓
Subtitles (SMIL, QT, SubRip)		✓

