

AG/EAF/EXMARaLDA etc.

Thomas Schmidt

SFB 538 ‚Multilingualism‘

University of Hamburg

thomas.schmidt@uni-hamburg.de

www.exmaralda.org

Time-Based Data Models / Interoperability for Audio/Video Transcription and Annotation

Thomas Schmidt

SFB 538 ‚Multilingualism‘

University of Hamburg

thomas.schmidt@uni-hamburg.de

www.exmaralda.org

Background

- Interoperability issues for a set of established transcription/annotation tools
- Different communities (Linguistics and LT) involved
- Joint work with Susan Duncan, Oliver Ehmer, Jeffrey Hoyt, Michael Kipp, Dan Loehr, Magnus Magnusson, Travis Rose, Han Sloetjes, Jan-Torsten Milde (tool developers and users)

References

- Thomas Schmidt, Susan Duncan, Oliver Ehmer, Jeffrey Hoyt, Michael Kipp, Dan Loehr, Magnus Magnusson, Travis Rose & Han Sloetjes (2009):
An exchange format for multimodal annotations.
To appear in: M. Kipp, J.-C. Martin, P. Paggio, and D. Heylen (eds.):
Multimodal corpora. Dordrecht: Springer.
(also in LREC Proceedings 2008)
- Thomas Schmidt (2005):
Time-based data models and the Text Encoding Initiative's guidelines for transcription of speech.
In: Arbeiten zur Mehrsprachigkeit (Working Papers in Multilingualism), Serie B (62). Hamburg.
Available from <http://www.exmaralda.org>

Tools (1): Anvil

The screenshot displays the Anvil 3.6 software interface, which is used for video analysis and gesture tracking. The interface is divided into several panels:

- Top Left Panel:** Contains the menu bar (File, Edit, View, Tools, Bookmarks, ?) and a toolbar with icons for file operations. Below the toolbar, it shows video specifications: "Loading video: IV50, 384x288, FrameRate=25", "Video frame rate: 25.0", "Audio format: LINEAR, 22050.0 Hz, 8-bit, Mono", and "Duration: 03:43:92 (5597 frames)". It also indicates "Open first player" and "wrote file lq1-7-reich.anvil".
- Top Middle Panel:** Displays the video frame being analyzed, showing a man in a suit and glasses with his hands raised in a gesture. The video is titled "Video: lq1-7-reich.avi".
- Top Right Panel:** Shows the "Track: gesture.phrase" information, including "Referenced track: gesture.phrase" and "Time: 03:07:63 - 03:10:27 (66 frames)". It lists attributes: "category: iconic", "iconic type: smash", "handedness: 2H", "cooc: Rivas", "function: emblematic", and "timing: direct". A comment field contains "compare with lq1-8 at 0:28".
- Bottom Panel:** The "Annotation: lq1-7-reich.anvil" section, which is a timeline view. It shows various tracks over time (03:06 to 03:11):
 - wave:** Audio waveform.
 - praat:** Praat spectrogram.
 - tri:** Text transcription: "übersetzen", "ich", "spüre", "den, poetischen", "Stil", "den, ich", "bei", "Rivas", "na", "nur, die, Bemühung".
 - tri2:** Empty track.
 - ling:** Empty track.
 - posture:** Tracks for "pose" (purple) and "shift" (grey).
 - phase:** Tracks for "phase" (pink), "prep" (red), "beats" (dark red), "hold" (light pink), "stroke" (red), and "hold" (light pink).
 - phrase:** Tracks for "phrase" (yellow), "metaphoric, heart, 2H" (orange), "iconic smash, 2H" (green), and "emblem, so-what, 2H" (red).
 - gesture:** Tracks for "gesture" (grey) and "compound" (grey).

Developer: Michael Kipp, DFKI Saarbrücken

Tools (2): ELAN

The screenshot displays the ELAN software interface for a file named 'Elan - r03_v20_s5.eaf'. The interface is divided into several sections:

- Video View:** Shows a scene with several people sitting on the floor in a traditional setting.
- Grid:** A table of annotations with columns for 'Nr', 'Annotation', 'Begin Time', 'End Time', and 'Duration'. The selected annotation (Nr 8) is highlighted in blue.
- Timeline:** A horizontal axis showing time from 00:01:00.000 to 00:01:09.000. A vertical red line indicates the current playback position at 00:01:01.115.
- Transcription Layers:** Multiple horizontal bars represent different layers of transcription, including:
 - Gloss1:** A red bar.
 - Gloss2:** A light blue bar containing the text: 'You have sung them, sing t | You sing those two part | You (P) ask him (Kp) | that thing, you sung those 2 parts of Myää, |
 - Kp:** A grey bar with 'mm' and ':éé?'
 - Plus:** A grey bar with 'Myää mbwaa dé a ngí | Myää kn:ää ló y:i a nyi cha |
 - K:** A grey bar with 'mu cha ngí dé, mēdē ngí d | Myää mbwaa dé... | u kwo péé, nyi u kwo pé | nyi u kwo p | mu tpile, mu Myää u mbwaa dé mu cha n |
 - Ricky:** A grey bar.
 - man outside vie:** A grey bar with '= K?'
 - child 1:** A grey bar.
 - loudspeaker:** A grey bar.

Developer: Han Sloetjes, MPI Nijmegen

Tools (3): EXMARaLDA Editor

EXMARaLDA Partitur-Editor 1.4.4 [T:\TP-Z2\DATEN\EXMARaLDA_DemoKorpus\PearStory\PearStory.exb]

File Edit View Transcription Tier Event Timeline Format Help

00:00 00:01 00:02 00:03

00:01.90 1.351 00:03.25

+ Add event... Append interval

	0 [00.0*]	1 [01.9]	2 [02.0]	3 [03.2]	4 [04.6]	5 [05.0]
		louder				
X [v]	So it starts out with: A	roo	ster crows.	{{1,4s}}	{{breathes in}}	And then you see ehm a
X [mv]	rHA on rKN, IHA on ISH	rHA up and to the right	rHA stays up		rHA back down rKN	
X [fm]						
X [mv]		emphasizes the crow				
Y [v]						

Done.

Transcription T:\TP-Z2\DATEN\EXMARaLDA_DemoKorpus\PearStory\PearStory.exb opened

Audio/Video panel

..DATEN\BEISPIELE\LYON\PearStory\pear.mov

Start

Position

Stop

0.0 3.2 5.1 43.8 43.8 sec

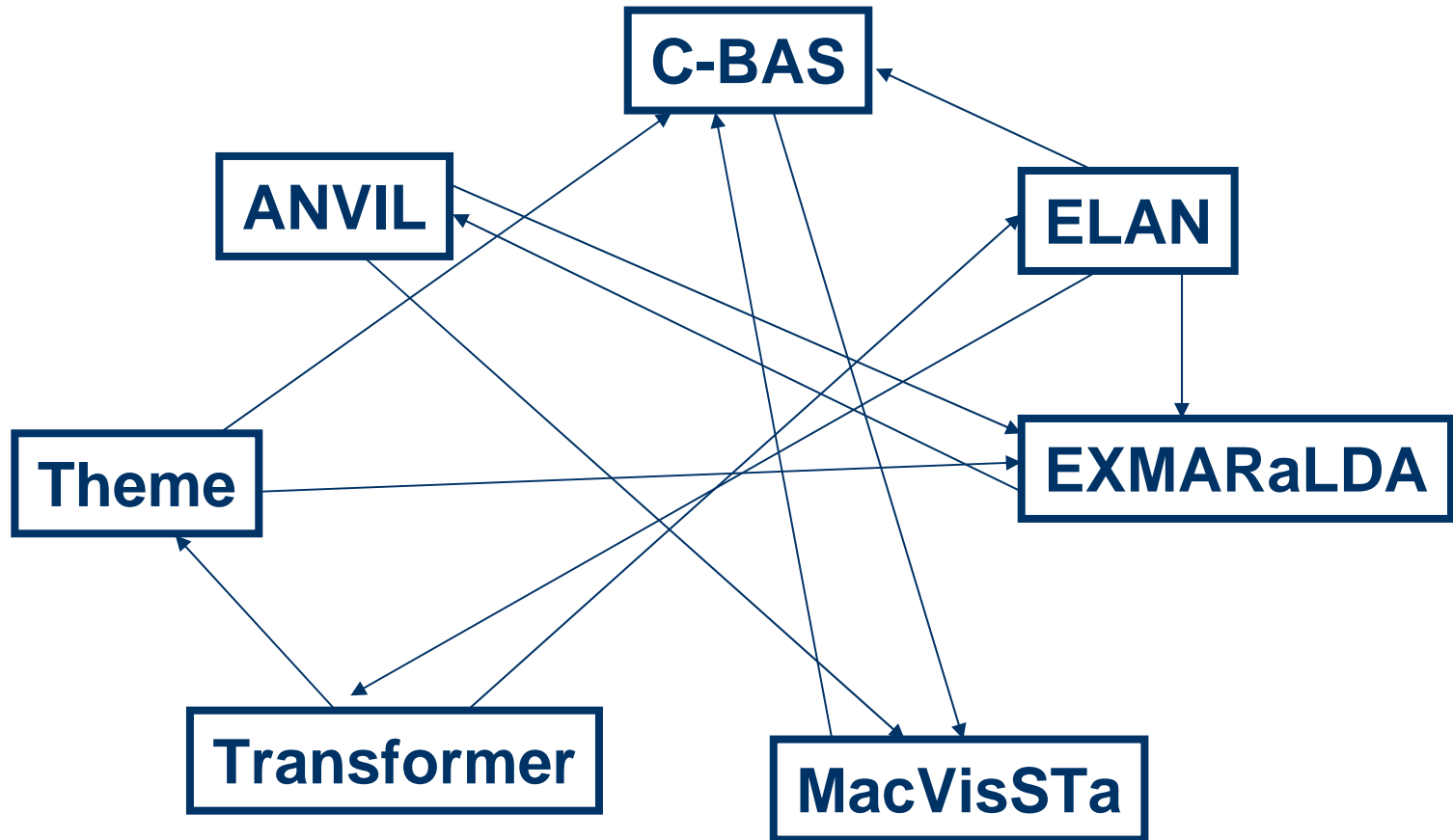
Playback halted

Developer: Thomas Schmidt, University of Hamburg

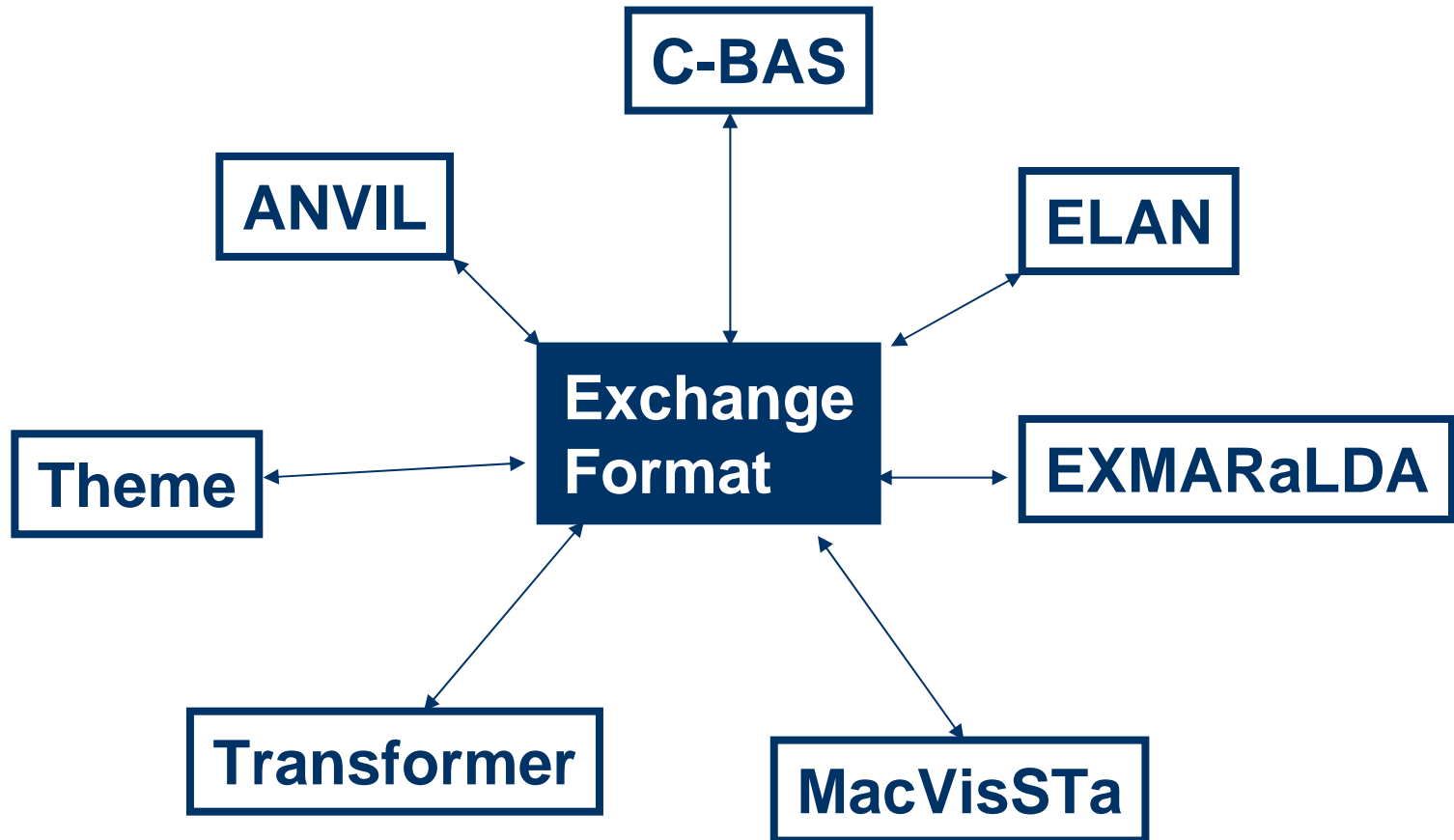
Multi-Layer Annotation Tools

- ANVIL, Praat, EXMARaLDA, ELAN, TASX- Annotator, MacVisSTA, Transformer, Theme, ...
- Developed for
 - Study of multimodal behaviour
 - Phonetic/Phonological analysis
 - Study of multi party conversation
 - Transcription + Glossing + Translation (field linguistics, endangered Languages)
- Also used in
 - Dialectology
 - Language acquisition studies
 - Multi-layer annotation of written texts

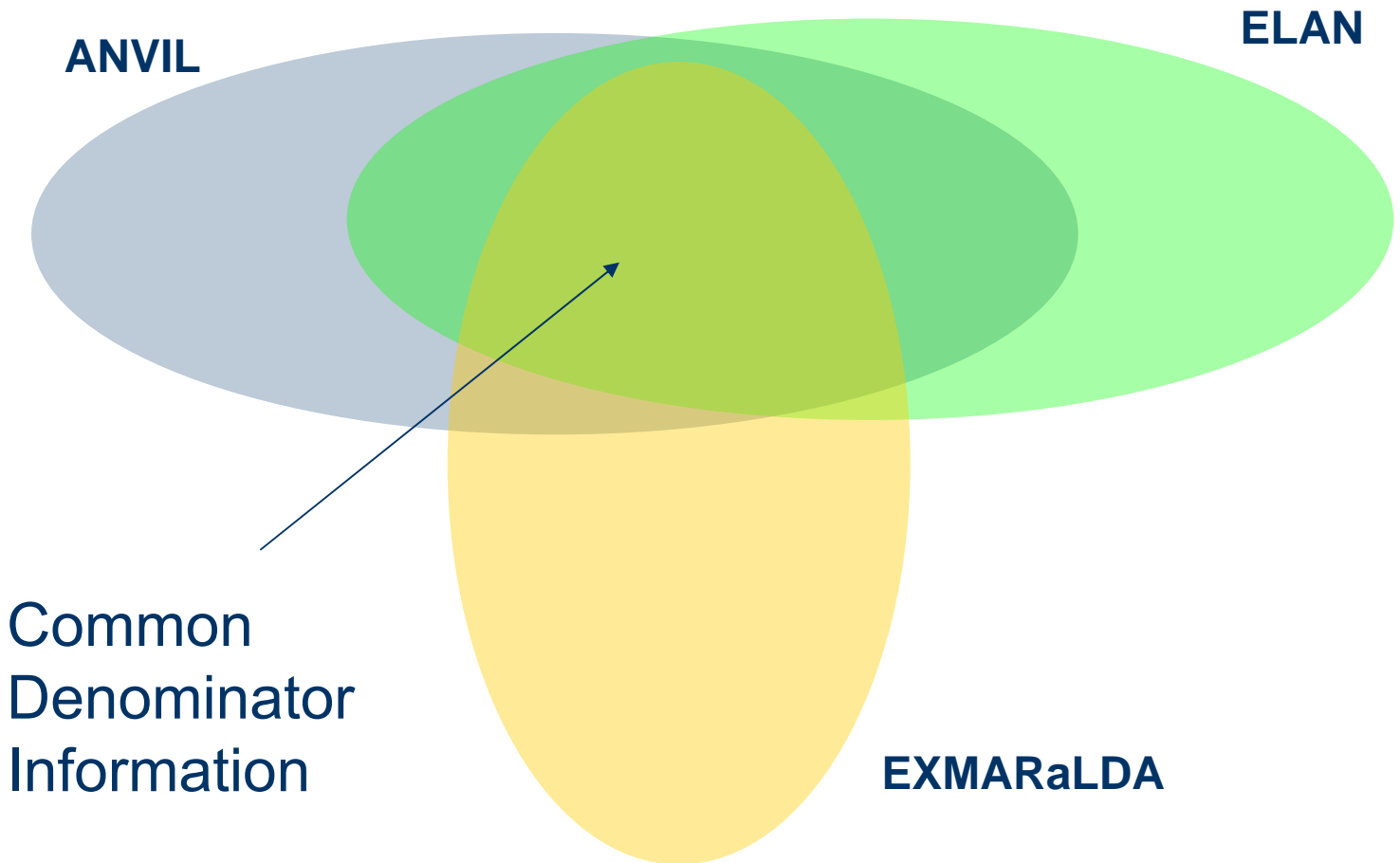
Interoperability



Interoperability



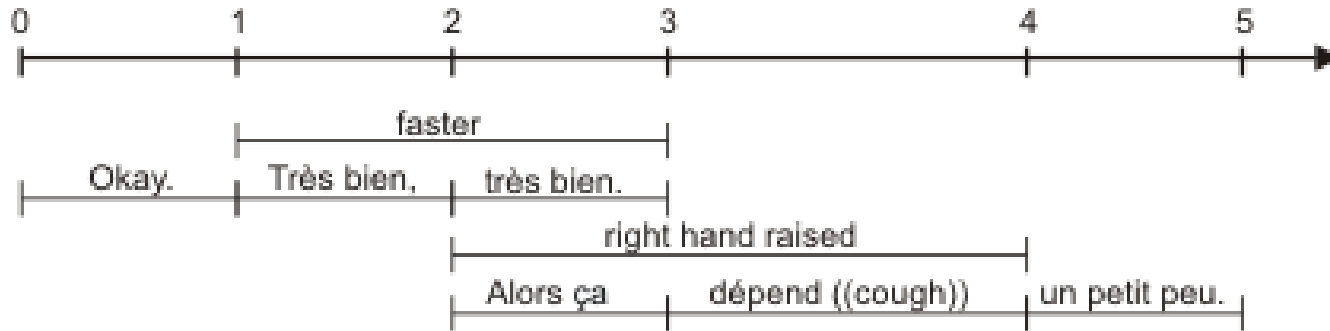
Data model comparison



Data model comparison

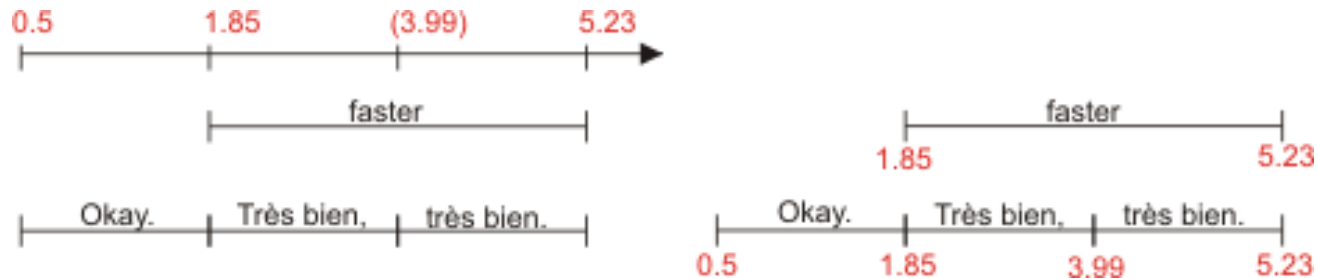
- Basic building blocks: Annotation tuples
 <start, end, label(s)>
 - Annotation Graphs as a general framework
 - AG's XML format as the base format
- Differences:
 - General organisation of basic building blocks into larger structural units
 - Semantic specifications and constraints on structural units

Tier-based vs. Non-tier-based



- Tiers = Partition of annotation tuples
 - Temporal overlap within a tier allowed?
- Construct partition from other information (e.g. categorisation of labels)

Implicit vs. explicit timeline



- Implicit timeline: annotation tuples refer directly to media times
 - Explicit timeline: annotation tuples refer to points in a timeline which can refer to media times
 - Relative and absolute ordering of timepoints
 - Timepoints without timestamps possible
- Interpolate timepoints without timestamps
- Construct explicit timeline (identical timestamps?)

Tier specifications

- Tier names (all)
- Speaker assignment (ELAN, EXMARaLDA)
- Tier types:
 - ANVIL: primary, singleton, span
 - ELAN/Transformer: time subdivision, included in, symbolic subdivision, symbolic association
 - EXMARaLDA: transcription, description, annotation

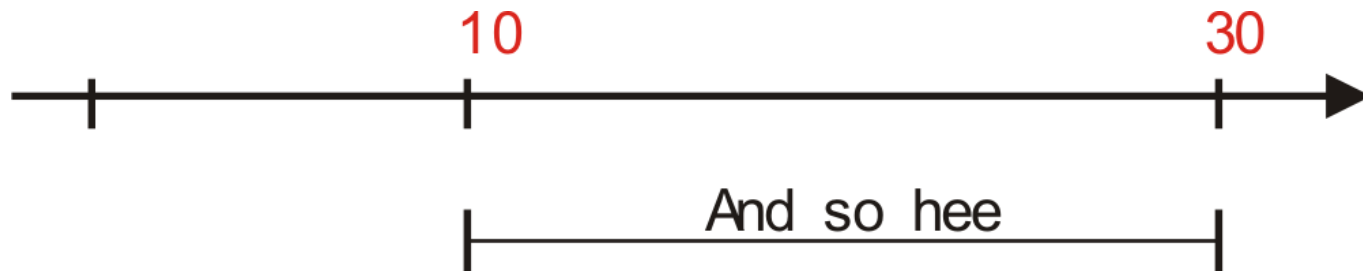
Tier relations and constraints

- Parent/Child relations → tier hierarchy
 - explicit in Anvil and ELAN
 - implicit in EXMARaLDA
- Other constraints arising from tier typing
- Restrictions on label content
 - part of the tools' format?

Exchange Format

- Lossless exchange of common denominator information
- Uniformly encode all information beyond the common denominator
- no lossless round-tripping, but...
- ... all available information captured and...
- ... lossless exchange in a chain of tools with increasingly complex data formats

Exchange Format



```
<Anchor id="T6" offset="10" unit="milliseconds"/>
<Anchor id="T7" offset="30" unit="milliseconds"/>
[...]
<Annotation type="TIE1" start="T6" end="T7">
  <Feature name="description">
    And so hee
  </Feature>
</Annotation>
```

Exchange Format

```
<MetadataElement name="Tier">
  <MetadataElement name="TierIdentifier">
    TIE1
  </MetadataElement>
</MetadataElement>
[...]
```

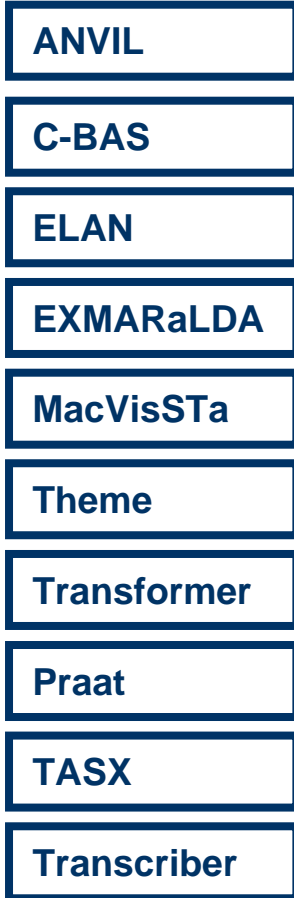
```
<Annotation type="TIE1" start=" T6" end=" T7">
```

```
<MetadataElement name="Tier">
  [...]
  <MetadataElement name="TierAttribute">
    <MetadataElement name="Source">
      EXMARaLDA
    </MetadataElement>
    <MetadataElement name="Name">
      speaker
    </MetadataElement>
    <MetadataElement name="Value">
      SPK0
    </MetadataElement>
  </MetadataElement>
</MetadataElement>
[...]
```

Tier definition:
Fixed metadata attribute
,TierIdentifier‘

Tier properties:
Fixed metadata triple
,Source‘
,Name‘
,Value‘

Conclusion



Results:

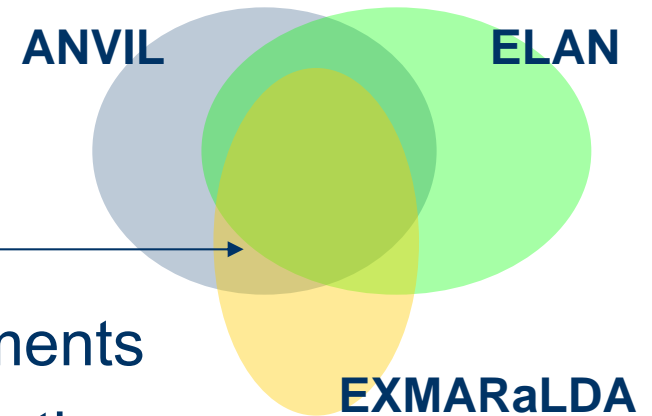
- Commonalities captured
- Differences (better) understood
- Basic interoperability established
- Link to generic framework established



Specific

Abstract

Open questions



- Partial correspondences?
 - Simple: e.g. speaker assignments
 - Complex: e.g. parent/child relations
- Exchange Format → Standard?
- Modifying/Assimilating tool formats?
- Time-based ↔ Hierarchy-based data models?

Theory? → Practice!

	ansatzweise angedeutet. Abkürzungen: LA= linker Arm, RA= rechter Arm, LH= linke Hand, RH= rechte Hand, KO= Kopf, OK= Oberkörper.
transcription-name	Rudi Völler Wutausbruch
Transkribent	Annette Schnieder
Transkription erstellt im	Januar 2004
Vorgeschichte	Das Fußball-EM-Qualifikationsspiels gegen Island endete 0:0. Vor der Live-Schaltung nach Reykjavik haben die Moderatoren der Sportsendung, Gerhard Dellling und Günter Netzer, die noch auf einem Monitor im Studio zu sehen sind, das Spiel bzw. die Leistung der deutschen Nationalmannschaft bewertet.

Speakers: WH; RV;

Location: Reykjavik, Island

Lokal ARD-Fernsehstudio

Start: 2003-09-06T21:30:00

Duration:

Recording (2.093 minutes): Rudi_Voeller_Wutausbruch.mp3

Recording (2.093 minutes): Rudi_Voeller_Wutausbruch.wav

Recording (2.093 minutes): Rudi_Voeller_Wutausbruch.mpg

Transcription Rudi

EXMARALDA: [Transcription] [Segmented]

Visualisation: [Utterances] [Words]

Export: [Transcription] [Segmented]

in Rudi

DA: [Transcription] [Segmented]

ion: [Partiture] [RTF] [PDF] [XML] [U

TEI] [AG] [EAF] [Praat] [Chat] [FOLKE

WH (Waldemar Hartmann)

Sex male

Ausbildung: Volontär bei der
beruflich "Schwäbischen Neuen Presse"

Ausbildung: Fachoberschulreife (mittlere
schulisch Reife) 1965

Ausgeübte Berufe Wirt in Augsburg Anfang der
70er Jahre.

Beruf Sportjournalist seit 1970

Familie Verheiratet (zum dritten Mal),
zwei Kinder.

Funktion Fernsehmoderator, Sport

Hobbies Golf, Tennis und Skilauf

Name Hartmann

Vorname Waldemar

Language: DEU

Status L1

Varietät Deutliche bayrische Aussprache
(sehr scharf gesprochenes s,
rollendes r, o statt a, usw.), teilweise
Dialekt (z. B. "ich war gestanden",
"Da müssen ma Tore schießen ...",
"Ja ich bin nit, du ich bin ja nit
beleidigt, Rudi ...").

In Communications: Rudi Völler: Wutausbruch;