

Computergestützte Erstellung und Auswertung von  
Korpora gesprochener Sprache mit  
**EXMARaLDA**

Thomas Schmidt,  
Projekt Z2 "Computergestützte Erfassungs- und  
Analysemethoden multilingualer Daten"  
SFB 538 'Mehrsprachigkeit'  
Universität Hamburg

# Gliederung

## Teil 1: Arbeiten mit EXMARaLDA

- Datenmodell
- Datenformat
- Software-Werkzeuge
- Nutzer

## Teil 2: Offene Fragen

- Datenmodelle für Mehrebenenannotation
- Software-Bausteine

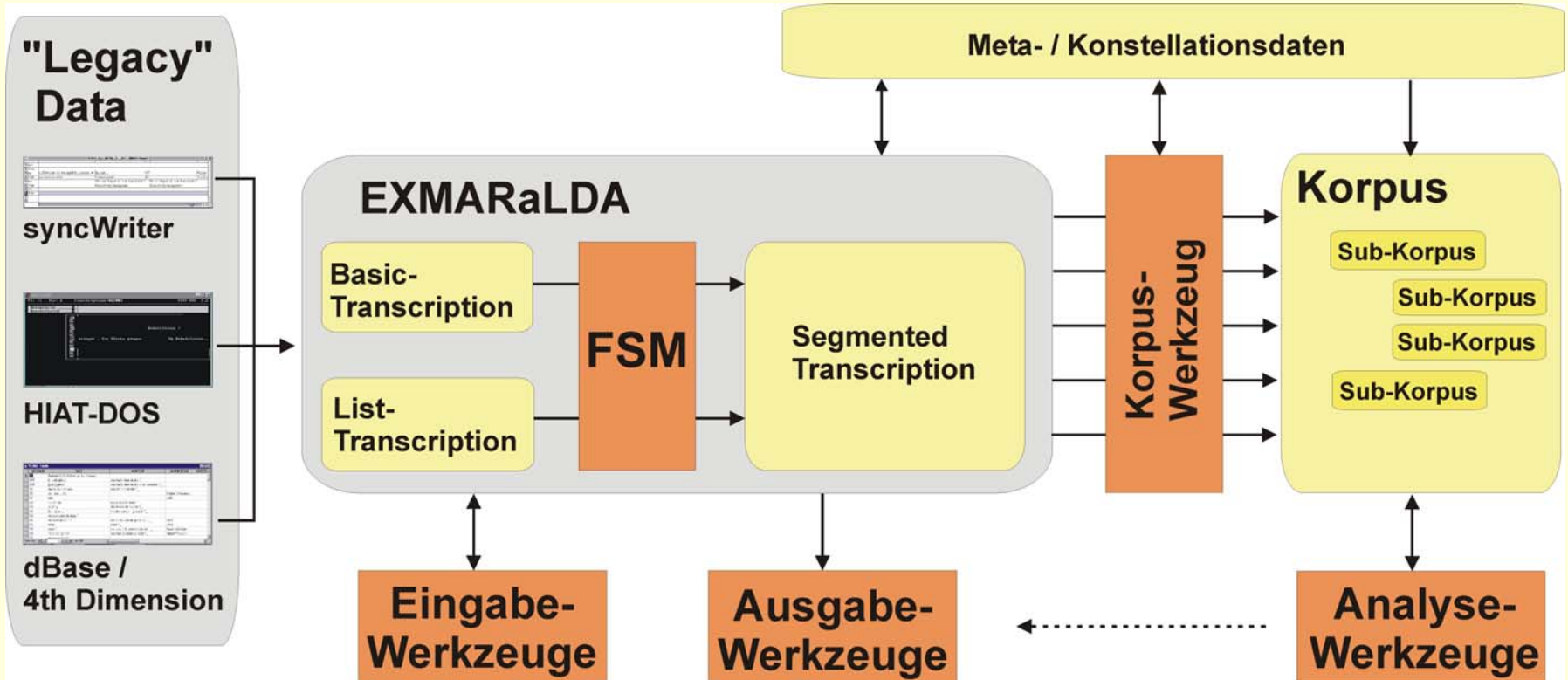
# Datenbank 'Mehrsprachigkeit'

Datenprojekt am SFB 538 'Mehrsprachigkeit'

Ziele:

- Rettung von Datenbeständen
- Archivierung
- Gemeinsamer Überbau für heterogene Datenbestände
- Zeitgemäße Transkriptionswerkzeuge
- Korpusverwaltung
- Auswertung

# Systemarchitektur



# EXMARaLDA

- Datenzentrierte Lösung
- Zeitbasiertes Datenmodell
- XML-basierte Datenformate
- Java-basierte Werkzeuge
- Interoperabilität
  - Import/Export: Praat, TASX, ELAN, AIF
  - Ausgabe: RTF, HTML, CSV, SVG

# Datenmodell

DS [sup]

*faster*

DS [v]

Okay. D'accord d'accord.

DS [en]

Okay. Agreed, agreed.

DS [nv]

*right hand raised*

FB [v]

Alors ça dépend ((cough)) un petit peu.

FB [en]

That depends then, a little bit.

FB [pho]

[étipø:]

# Datenmodell

DS [sup]	<i>faster</i>
DS [v]	Okay. D'accord d'accord.
DS [en]	Okay. Agreed, agreed.
DS [nv]	<i>right hand raised</i>
FB [v]	Alors ça dépend ((cough)) un petit peu.
FB [en]	That depends then, a little bit.
FB [pho]	[étipø:]

## Strukturelle Beziehungen:

1. Zeitliches Nacheinander

# Datenmodell

DS [sup]	<i>faster</i>	
DS [v]	Okay. D'accord d'accord.	
DS [en]	Okay. Agreed, agreed.	
DS [nv]		<i>right hand raised</i>
FB [v]		Alors ça dépend ((cough)) un petit peu.
FB [en]		That depends then, a little bit.
FB [pho]		[étipø:]

## Strukturelle Beziehungen:

1. Zeitliches Nacheinander
2. Gleichzeitigkeit

# Datenmodell

DS [sup]	<i>faster</i>	
DS [v]	Okay. D'accord d'accord.	
DS [en]	Okay. Agreed, agreed.	
DS [nv]	<i>right hand raised</i>	
FB [v]	Alors ça dépend ((cough)) un petit peu.	
FB [en]	That depends then, a little bit.	
FB [pho]		[étipø:]

## Strukturelle Beziehungen:

1. Zeitliches Nacheinander
2. Gleichzeitigkeit
3. Äquivalenz ('flache' Annotation)

# Datenmodell

DS [sup]

*faster*

DS [v]

Okay. D'accord d'accord.

DS [en]

Okay. Agreed, agreed.

DS [nv]

*right hand raised*

FB [v]

Alors ça dépend ((cough)) un petit peu.

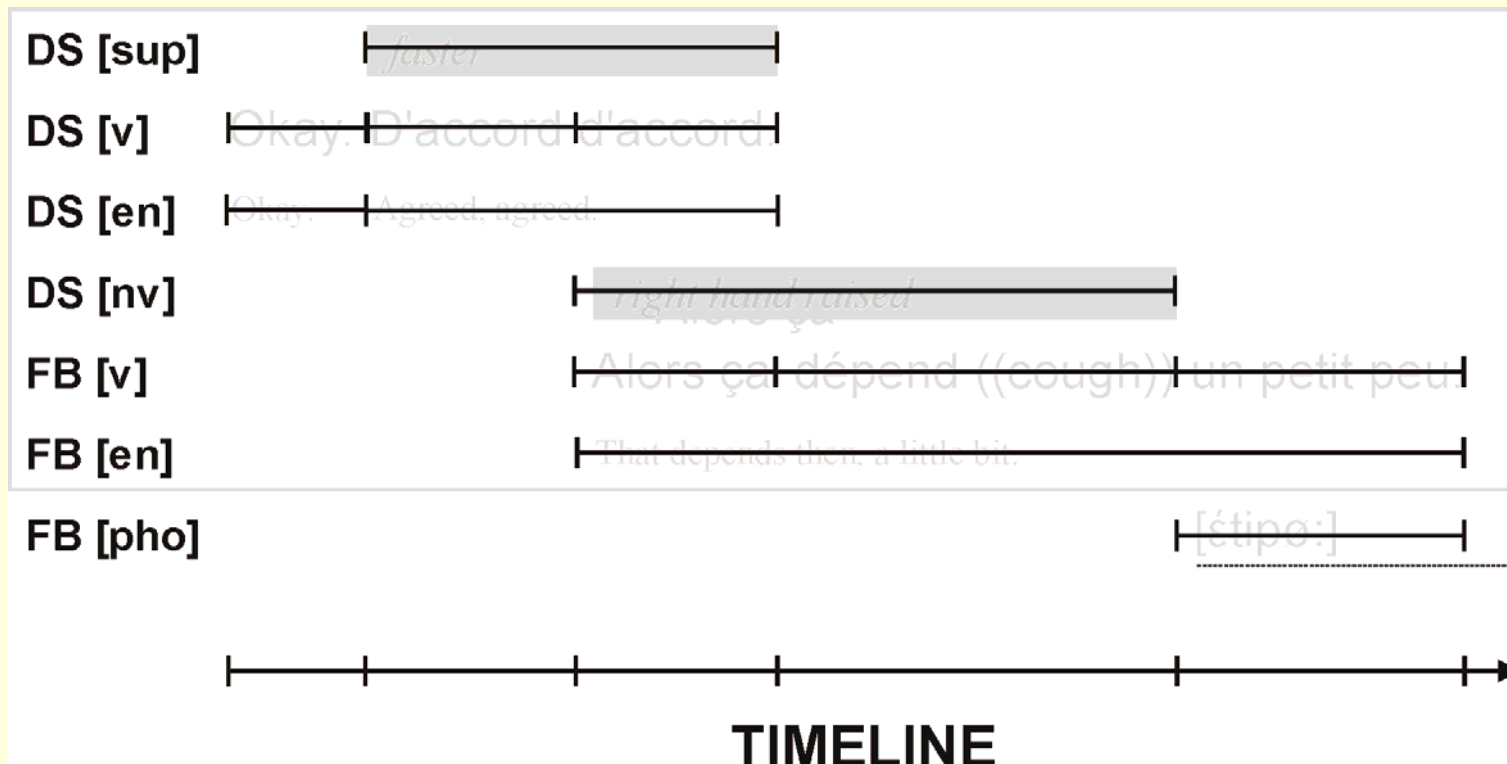
FB [en]

That depends then, a little bit.

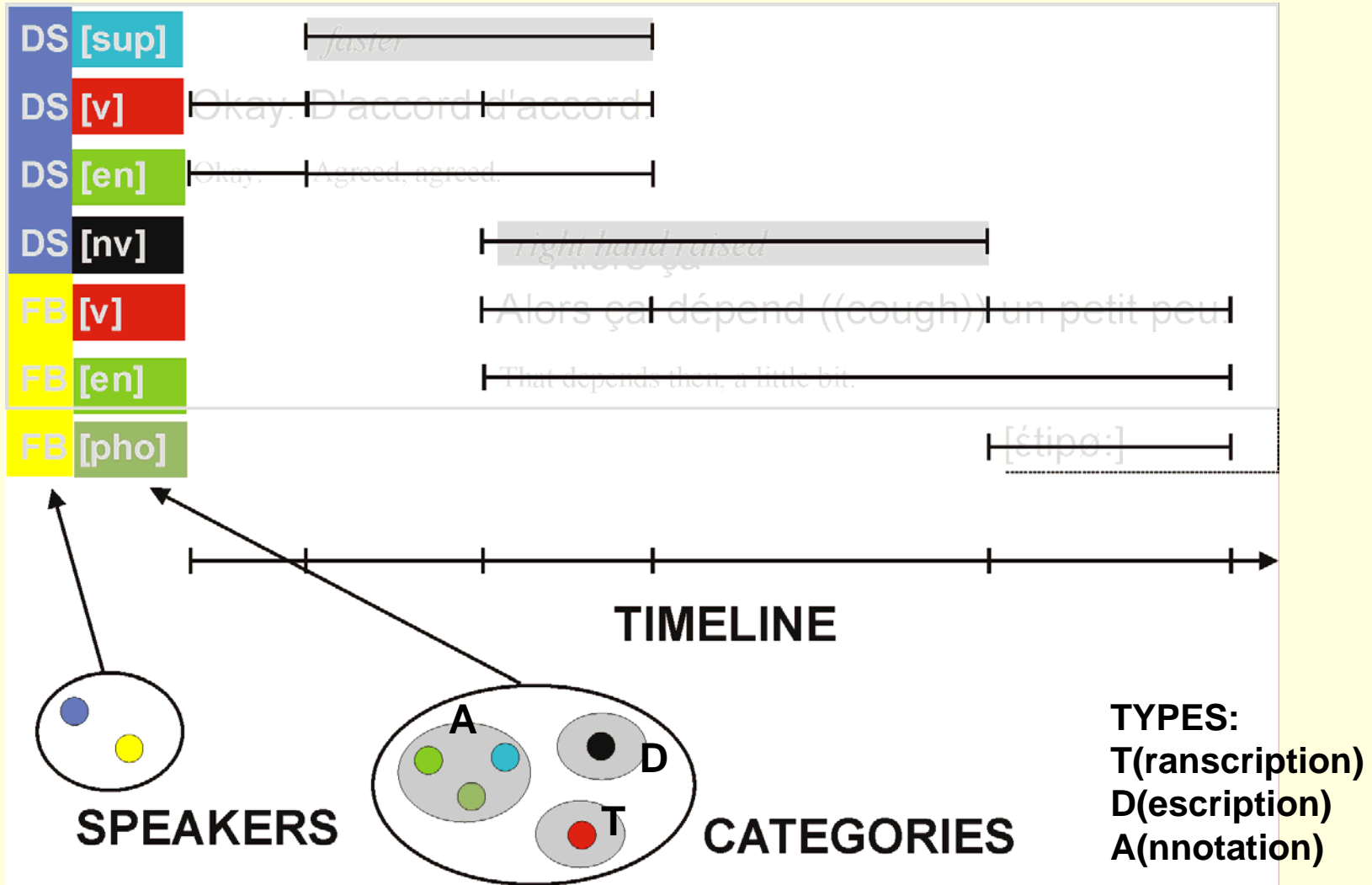
FB [pho]

[étipø:]

# "Single timeline, multiple tiers"



# "Single timeline, multiple tiers"



# Datenformat

```
<speakertable>
  <speaker id="SPK_KLA">
    <abbreviation>KLA</abbreviation>
    <sex value="m"/>
  </speaker>
  <speaker id="SPK_SCHNEIDER">
    <abbreviation>SCHNEIDER</abbreviation>
    <sex value="m"/>
  </speaker>
</speakertable>
```

```
<common-timeline>
  <tli id="ts1" time="0.855" bookmark="START"/>
  <tli id="ts3" time="1.4999923308692669"/>
  <tli id="T19" time="3.1999741901195726"/>
  <tli id="ts4" time="3.835"/>
  <tli id="ts7" time="4.63"/>
  <tli id="ts9" time="5.779"/>
```

```
<tier id="KLA" speaker="SPK_KLA" category="v" type="t" display-name="KLA [v]">
  <event start="ts1" end="ts3">Guten Tag! </event>
  <event start="ts4" end="ts7">Jaa. </event>
  <event start="ts7" end="ts9">Dann woll'n wer mal schau. </event>
```

```
<tier id="SCHNEIDER" speaker="SPK_SCHNEIDER" category="v" type="t" display-name="SCHNEIDER [v]">
  <event start="ts3" end="T19">Guten Tach, mein Name is Schneider. </event>
  <event start="T19" end="ts4">Ich such ne Ar</event>
  <event start="ts4" end="ts7">beit. </event>
```

# Datenformat

```
<speakertable>
  <speaker id="SPK_KLA">
    <abbreviation>KLA</abbreviation>
    <sex value="m"/>
  </speaker>
  <speaker id="SPK_SCHNEIDER">
    <abbreviation>SCHNEIDER</abbreviation>
    <sex value="m"/>
  </speaker>
</speakertable>
```

```
<common-timeline>
  <tli id="ts1" time="0.855" bookmark="START"/>
  <tli id="ts3" time="1.4999923308692669"/>
  <tli id="T19" time="3.1999741901195726"/>
  <tli id="ts4" time="3.835"/>
  <tli id="ts7" time="4.63"/>
  <tli id="ts9" time="5.779"/>
```

```
<tier id="KLA" speaker="SPK_KLA" category="v" type="t" display-name="KLA [v]">
  <event start="ts1" end="ts3">Guten Tag! </event>
  <event start="ts4" end="ts7">Jaa. </event>
  <event start="ts7" end="ts9">Dann woll'n wer mal schau'n. </event>
```

```
<tier id="SCHNEIDER" speaker="SPK_SCHNEIDER" category="v" type="t" display-name="SCHNEIDER [v]">
  <event start="ts3" end="T19">Guten Tach, mein Name is Schneider. </event>
  <event start="T19" end="ts4">Ich such ne Ar</event>
  <event start="ts4" end="ts7">beit. </event>
```

# Werkzeuge

- Partitur-Editor
  - Eingabe, Ausgabe von Transkriptionen
- Corpus-Manager (CoMa)
  - Verwaltung, Durchsuchen von Korpus-Metadaten
- Suchwerkzeug (Zecke, Squirrel)
  - Suche, Auswertung von Transkriptionskorpora

# Partitur-Editor

EXMARALDA Partitur-Editor 1.3.2. PV1 [I:\Exmaralda\Annette\LYON\Pear Story\BasicTransPear.xml]

Datei Edit View Tier Event Timeline Format Segmentation Help

	0 [0.]	1 [1.9]	2 [1.9]	3 [3.2]	4	5 [9.2]	6 [10.4]	7 [11.1]
		louder						
X [v]	So it starts out with: A	roo	ster crows	. ((1,3s)) ((takes breath 0,5s))	And then you see ehm a maan in maybe	his fifties. ((takes breath, 0,5s))	And so hee	climbs
X [mv1]	rHA on rKN, lHA on ISH	rHA up and to the right	rHA stays up	rHA back down rKN, moves lHA on ISH	rHA to the right	bHA up, rHA rubs NO		
X [mv2]			HE nods once					
X [mv3]		emphasizes the crow				approximately fifty		
Y [v]								
Vid	Shot-1	Shot-2				Shot-3		

Done.

Audio/Video panel

...maralda\Annette\LYON\Pear Story\pear.mov

Start

Position

Stop

0.0 15.1 15.1 43.8 43.8 sec

Start time changed: 15.067463062773067sec.

Keyboard

International Phonetic Alphabet

Open...

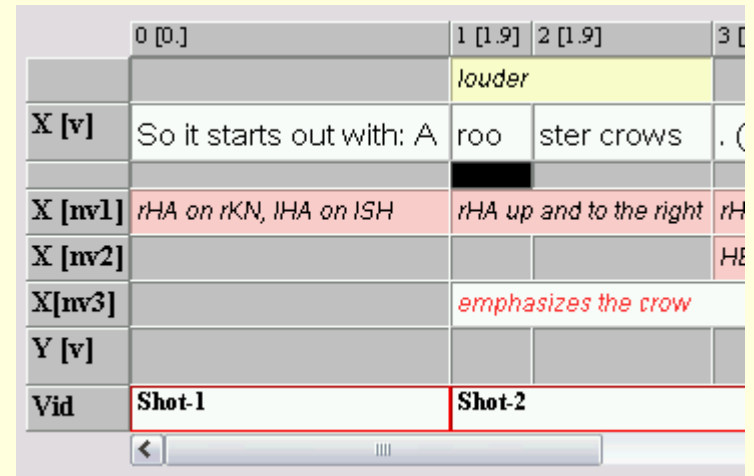
e a o b c d e s e e c  
 z 3 e j g g e y v u h h  
 i l I t t l B w u m n n  
 N e e o p j j l r r r r  
 R v s f f l j t t t u u  
 A M A Y z z z a ? s j c  
 o B a g H j k L d ? s d z  
 æ œ ɛ ɣ ɰ

# Partitur-Editor

## Transkribieren

### 1. direkt im Partitur-Editor

- Sprechertabelle bearbeiten
- Partitur L&F
- Spuren anlegen, umstellen, löschen, aus-/einblenden
- Ereignisse beschreiben



	0 [0.]	1 [1.9]	2 [1.9]	3 [1.9]
		<i>louder</i>		
X [v]	So it starts out with: A	roo	ster crows	. (
X [nv1]	<i>rHA on rKN, lHA on ISH</i>	<i>rHA up and to the right</i>	<i>rH</i>	
X [nv2]				<i>H</i>
X [nv3]			<i>emphasizes the crow</i>	
Y [v]				
Vid	Shot-1	Shot-2		

### 2. mit anderen Software-Werkzeugen (Import-Funktionen)

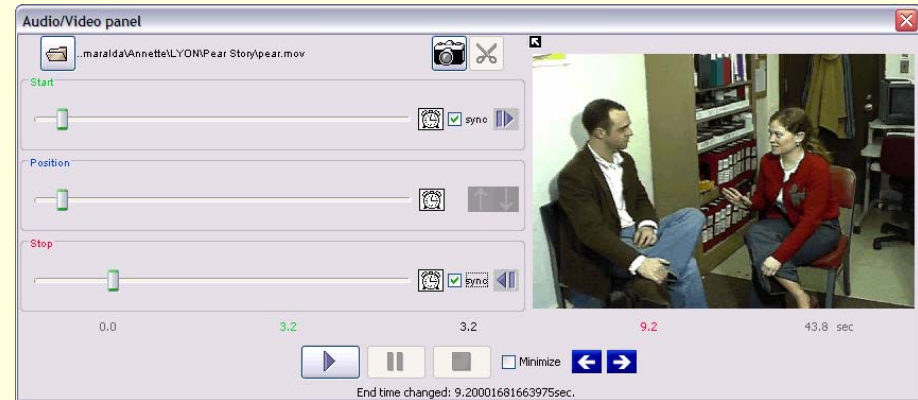
- mit Praat
- mit TASX oder ELAN
- in einer Textdatei ("Simple Exmaralda"-Konventionen)

# Partitur-Editor

## Synchronisieren

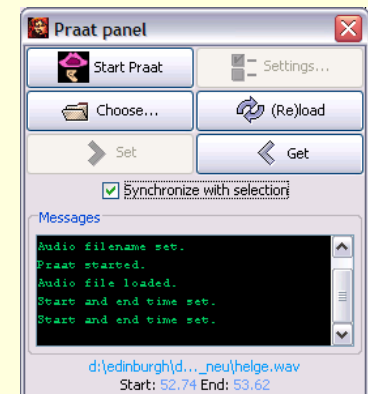
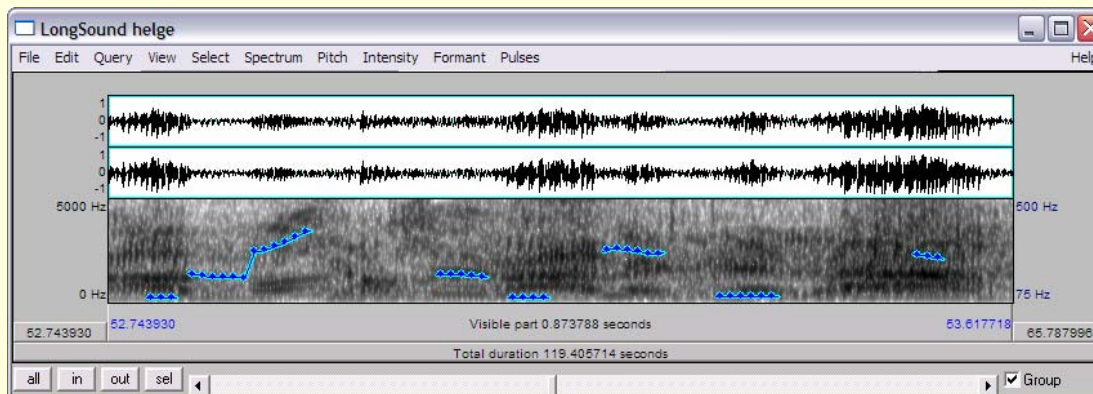
Über das Audio/Video-Panel

- Audio/Video über JMF
- wav, mp3, avi, mpeg



Über das Praat-Panel

- Audio mit diversen Visualisierungen
- Oszillogramm, Spektrogramm, etc.



# Partitur-Editor

## Verknüpfen

einzelne Ereignisse...  
...mit Audio-/Video-Ausschnitten  
...mit Bildern

Realisierung in der Ausgabe als  
Hyperlinks



# Partitur-Editor

## Ausgabe

Als umgebrochene Partitur in  
RTF, HTML, SVG,  
Drucker (PDF)

[1]	DS [sup]	<i>faster</i>
	DS [v]	Okay. D'accord d'accord.
	DS [en]	Okay. Agreed, agreed.
	DS [nv]	<i>right hand raised</i>
	FB [v]	Alors ça dépend
	FB [en]	That depends, then, a little bit
[2]	DS [nv]	
	FB [v]	((cough)) un petit peu.
	FB [en]	
	FB [pho]	[ɛ̃tipø:]

Über XSLT-Transformationen

Zusätzliche Ausgabemöglichkeiten durch → Segmentierung

# Partitur-Editor

## Segmentierung

≈ Tokenization: Unterteilung in Äußerungen, Wörter etc.  
 Zeitliche Struktur ⇔ Sprachliche Struktur  
 Automatisch mit Hilfe von Finite State Machines (FSM)  
 Abhängig von Transkriptionskonventionen  
 FSMs für HIAT, GAT, DIDA, CHAT

	0	1	2	3	4	5
A [v]	Guten Tag.		Wie • geht es Dir?	Alles	((hustet)) klar?	
B [v]		Hallo.		Gut.		

<b>Turn</b>																
<b>Äußerung</b>										<b>Äußerung</b>						
Wort	IP	Non-Pho	IP	Wort	IP	Wort	IP	Wort	IP	Wort	IP	Non-Pho	IP	Wort	IP	
Wie		•		geht		es		Dir	?	Alles		((	hustet	))	klar	?

# Partitur-Editor

## Segmentierung

### Weitere Ausgabeformen

- Äußerungslisten
- Wortlisten

### Basis für weitere Auswertung

- Quantifizierung
- Suche



# Corpus-Manager

Coma2 | testkorpus.xml | Corpus: corpusBasket from ENDFAS

Datei Werkzeuge Hilfe

Korpus Kommunikation Person Korpus-Korb Suchen Einstellungen

Angezeigter Key aus Description

XPath anzeigen Desc.-Key anzeigen

Person

Sigle	Var
Bil	Bilgi
Sin	Sinem
Fig	Figen
Sel	Selim
Selm	Selma
Ert	Ertan
Bur	Burçin
Vat	Vater
Mut	Mutter
Dan	Daniela
Azi	Azim
Mut	Mutter
Alt	Altan
Mut	Mutter
Gök	Gökkan
Seh	Şehmuz

Aktion

+ Person hinzufügen

- Person löschen

Alle Transkriptionen in den Korb

Daten

Schlüssel	Wert
Sigle	Fig
KnownHuman	true
Pseudo	Figen
Sex	female
Location (D)	Location bearbeiten
+Location	Neue Location hinzufügen
Description	Description bearbeiten
Language (TRK TR)	Language bearbeiten
Language (GER DE)	Language bearbeiten
+Language	Language hinzufügen

zugeordnete Kommunikationen

S	Name	Var
<input type="checkbox"/>	0610	EFE04 tk - Maulwurf
<input type="checkbox"/>	0675	EFE04 dt - Maulwurf
<input type="checkbox"/>	0672	EFE04 dt - Maulwurf
<input type="checkbox"/>	1577	EFE04 tk - Maulwurf
<input type="checkbox"/>	0549	EFE04 dt - Maulwurf
<input type="checkbox"/>	0535	EFE04 dt - Maulwurf
<input type="checkbox"/>	0572	EFE04 dt - Maulwurf
<input type="checkbox"/>	1580	EFE04 tk - Maulwurf
<input type="checkbox"/>	0715	EFE04 tk - Maulwurf
<input checked="" type="checkbox"/>	0515	EFE04 tk - Maulwurf
<input type="checkbox"/>	0686	EFE04 dt - Maulwurf
<input type="checkbox"/>	1579	EFE04 tk - Maulwurf
<input type="checkbox"/>	0505	EFE04 tk - Maulwurf

Schlüssel Wert

SprecherNr.	107
Eigenschaft	Nachbarskind
Vorname	Filiz
Nachname	Taşdan
zur Familie	Taşkan

neue Zeile OK Abbrechen Description löschen

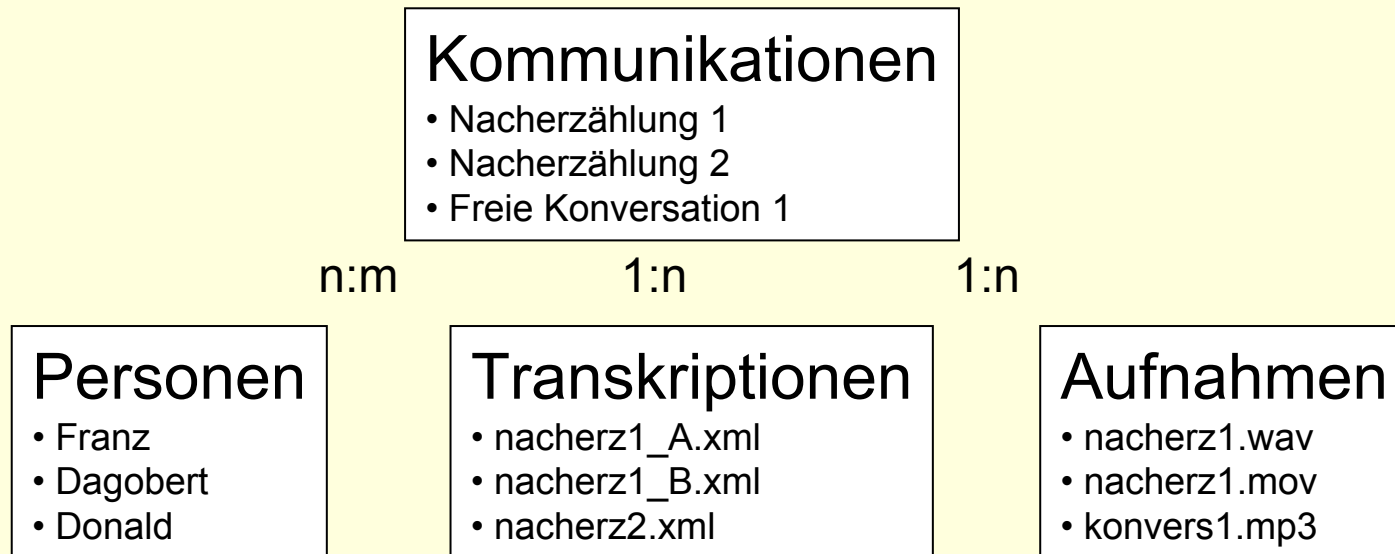
# Corpus-Manager

Eingabe von Meta-Daten zu

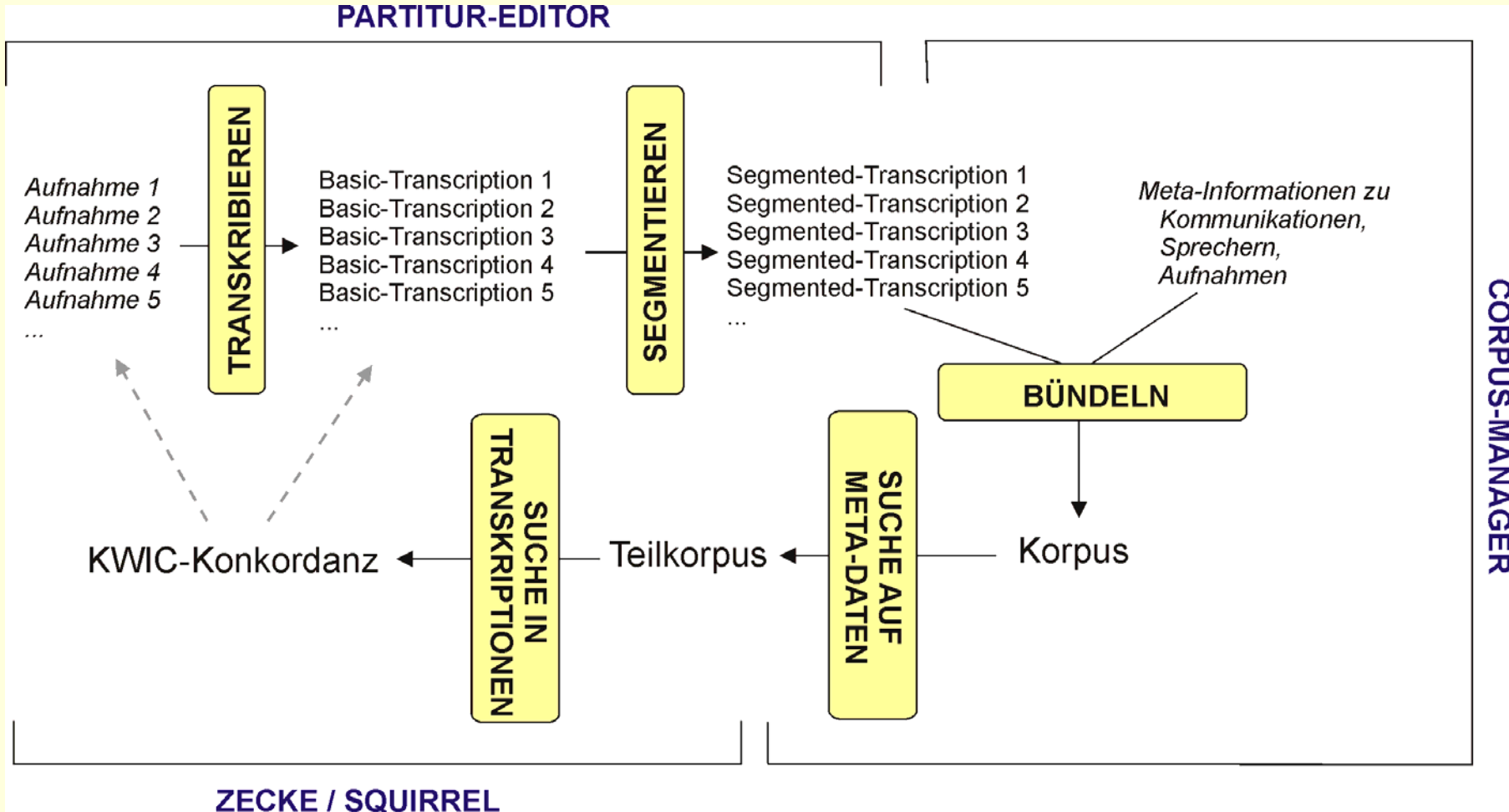
- Kommunikationen (z.B. Ort, Zeit)
- Personen (z.B. Name, Alter)

Zuordnen von Transkriptionen, Aufnahmen und Sprechern zu Kommunikationen

Suche auf Meta-Daten → Teilkorpora



# Workflow



# Suchwerkzeug

**ZECKE (Ziemlich Einfaches Konkordanztool für Exmaralda)**

**Search Expression:** (b[wW]iesolb)|(b[wW]eshalb)|(b[wW]arumb)

**Corpus**

- D:\Edinburgh\D\AAA\_Beispiele\Arbeitsamt\A
- D:\Edinburgh\D\AAA\_Beispiele\Helge\_neu\H
- D:\Edinburgh\D\AAA\_Beispiele\DIDA\RV\_se

No	Transcript	Speaker	Left context	Match	Right context
0	Tropfsteinhöhle	KLA		Höh, warum	lacht der denn jetzt?
1	Tropfsteinhöhle	KLA		Wie, warum	denn so abrupt?
2	Tropfsteinhöhle	KLA		Wieso	denn? Bleib doch noch n bißchen hier, k
3	Rudi Völler erregt	GN	t zum schluss jetzt auch noch ke"gel-*	warum	sagen sie nicht einfach das war n schön
4	Rudi Völler erregt	WH	* >ij/< ich * kann jetzt nicht verstehen	warum	die schä"rfe reinkommt   warum die (ich)
5	Rudi Völler erregt	WH	verstehen warum die schä"rfe reinkommt	warum	die (ich)
6	Rudi Völler erregt	WH	ämlich schon da   * wo wir waren   * >äh<	warum	so"ln wir nicht zu dem du kommen zu/
7	Rudi Völler erregt	WH	chon heute auf dem feld gewe"sen   * und	warum	haben sie=s heu"te nicht gemacht   *
8	Rudi Völler erregt	RV		wieso	müssen wir denn die mannschaften klar b
9	Rudi Völler erregt	RV	n die mannschaften klar behe"rrschen-*	wieso	wir spielen doch- * die isländer sind t
10	Rudi Völler erregt	RV	aben→ des ich kann das nicht verste"hen	warum	* ←island is→ * tabellenführer gewesen

**Partitur**

	60 [59.3]	61	62	63	64	65 [62.6]	66	67
<b>KLA [v]</b>	Ja, Zimmer vierzehn.		Ja, aber, aber...		Wie, warum denn so abrupt?			W
<b>KLA [en]</b>	Yes, room fourteen.		But, but...		Why so abruptly?			Bc
<b>KLA [fr]</b>	Oui, salle quatorze.		Mais, mais...		Pourquoi si brusquement?			M6
<b>ERW [v]</b>	...vierzehn.	Da war ich noch nicht drin.		Tschüß.		Ja, nee, och, ähm, ja hab ich aber jetzt keine Lust		t
<b>ERW [en]</b>	....	Never been there.		Bye.		Yes, no, eh, no, I don't feel like it anymore.		

100 % 0.0 seconds / 11 results.

**Audio/Video panel**

0.0 23.9 23.9 57.5 119.4 sec

Minimize

Playback stopped.

# Suchwerkzeug

## 1. ZECKE

- Prototyp (Ziemlich EinfaChes Konkordanzwerkzeug für Exmaralda)
- "Volltextsuche" in Transkriptionen über reguläre Ausdrücke
- KWIC-Konkordanz → Transkription (Partitur, Meta-Daten) → Aufnahme

## 2. SQUIRREL

- Search and Query InstRument foR ExmaraLda
- "Volltextsuche" in Transkriptionen über reguläre Ausdrücke
- Suche auf Annotiertem (nicht Transkribierten)
- XPath-Suchausdrücke
- KWIC-Konkordanz → Transkription (Partitur, Meta-Daten) → Aufnahme

# EXMARaLDA: Nutzer

## Diskursforschung / Spracherwerbsforschung / Mehrsprachigkeitsforschung

Am SFB 538 'Mehrsprachigkeit'

- 5 Projekte in der laufenden + 3 in der kommenden Phase

Am Institut für Deutsche Sprache (IDS) in Mannheim

In mehreren DFG-Projekten (Uni Dortmund, Uni Mannheim, Uni Flensburg)

In mehreren Dissertationsprojekten

In Lehrveranstaltungen

- an zahlreichen deutschen Universitäten
- vereinzelt im Ausland (Schweiz, USA, UK, Frankreich, Italien)

## Erziehungs- / Kommunikationsforschung

Am GK 'Bildungsgangforschung' in Hamburg

In der Mathematik-Didaktik in Köln, Kassel, Agder University (N)

## Annotation geschriebener Sprache

Am SFB 632 'Informationsstrukturen' in Potsdam

Am Lehrstuhl 'Korpuslinguistik' der HU Berlin

## Multimodale Kommunikation

# EXMARaLDA: Korpora

## SFB-Korpora

- Spracherwerb deutsch-türkisch bilingualer Kinder
- Spracherwerb deutsch-spanisch bilingualer Kinder
- Skandinavische (dänisch/schwedisch/norwegische) Semikommunikation
- Japanisch/Deutsche Expertendiskussion
- Dolmetschen (Deutsch/Portugiesisch/Türkisch) im Krankenhaus
- Dänisch/Färöische Zweisprachigkeit
- Spanisch/Katalanische Zweisprachigkeit
- Simultandolmetschen Portugiesisch/Deutsch

Korpus "Divergierender bilingualer Sprachgebrauch bei Jugendlichen",  
Uni Flensburg

Korpus "Frühe Mehrsprachigkeit", Uni Mannheim

Korpus "Interviews mit Israelis deutscher Primärsprache", Uni Dortmund

Korpora am IDS?

...?

# Teil 2: Offene Fragen

# Varianten von "Single timeline, multiple tiers"

EXMARaLDA (basic-transcription)	TASX
<pre> &lt;tier speaker="TOM" cat="v"&gt;   &lt;event start="T1" end="T2"&gt;     Tut mir   &lt;/event&gt;   &lt;event start="T2" end="T3"&gt;     leid.   &lt;/event&gt; &lt;/tier&gt; </pre>	<pre> &lt;layer l-id" TOM [v]"&gt;   &lt;event start="1.0" end="2.0"&gt;     Tut mir   &lt;/event&gt;   &lt;event start="2.0" end="3.0"&gt;     leid.   &lt;/event&gt; &lt;/tier&gt; </pre>
PRAAT	ELAN (Eudico Abstract Corpus Model)
<pre> item [3]:   class = "IntervalTier"   name = "TOM [v]"   xmin = 0.0   xmax = 3.0   intervals [1]:     xmin = 1.0     xmax = 2.0     text = "Tut mir "   intervals [2]:     xmin = 2.0     xmax = 3.0     text = "leid. " </pre>	<pre> &lt;tier participant="TOM" [...]&gt;   &lt;annotation&gt;     &lt;[...] timeslot_ref_1="T1"       timeslot_ref_2="T2"&gt;       &lt;value&gt;Tut mir &lt;/value&gt;     &lt;/[...]&gt;   &lt;/annotation&gt;   &lt;annotation&gt;     &lt;[...] timeslot_ref_1="T2"       timeslot_ref_2="T3"&gt;       &lt;value&gt;leid. &lt;/value&gt;     &lt;/[...]&gt; &lt;/tier&gt; </pre>

## Varianten von "Single timeline, multiple tiers"

	<b>EXMARaLDA</b>	<b>TASX</b>	<b>Praat</b>	<b>ELAN</b>
<b>Tier classification</b>	Types, Categories and speakers	Tier names	Tier names	Stereotypes, Linguistic Types and Participants
<b>Timeline</b>	Relative and/or absolute	Absolute	Absolute	Relative and/or absolute
<b>Overlap within tier</b>	No	Yes	No	Yes („Bulldozer mode“)
<b>Link to media</b>	Optional (Audio only)	Required (Video and Audio)	Required (Audio only)	Optional (Video and Audio)
<b>Extensions</b>	<b>Segmented Transcription</b>	TASX Level 2	None	Symbolic subdivisions, symbolic associations

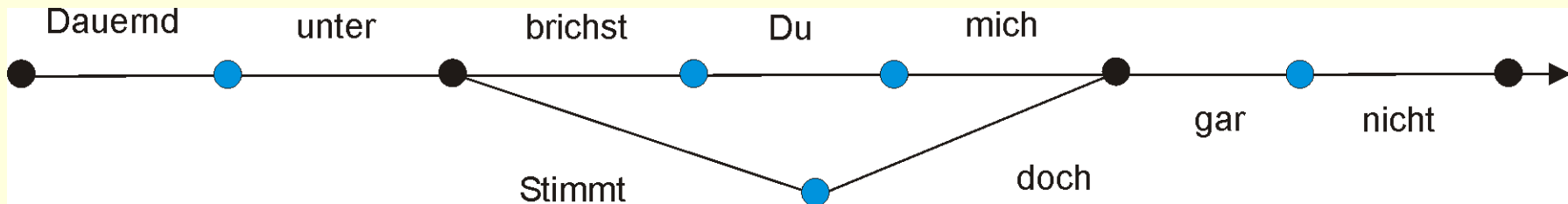
# EXMARaLDA Basic-Transcription

	T0	T1	T2	T4
A	Dauernd unter	brichst Du mich!		
B		Stimmt doch	gar nicht.	

```
<tier id="TIE0" speaker="SPK0" category="v" type="t" display-name="A">  
  <event start="T0" end="T1">Dauernd unter</event>  
  <event start="T1" end="T2">brichst Du mich! </event>  
</tier>  
<tier id="TIE1" speaker="SPK1" category="v" type="t" display-name="B">  
  <event start="T1" end="T2">Stimmt doch </event>  
  <event start="T2" end="T4">gar nicht. </event>  
</tier>
```

# Verzweigende Zeitachse / partiell geordnete Graphen

	T0	T1	T2	T4
A	Dauernd unter	brichst Du mich!		
B		Stimmt doch	gar nicht.	



# Verzweigende Zeitachse / partiell geordnete Graphen

	T0	T1	T2	T4
A	Dauernd unter	brichst Du mich!		
B		Stimmt doch	gar nicht.	

```
<Zeitachse>
  <t id="T0"/>
  <t id="Tx"/>
  <t id="T1"/>
  <Verzweigung>
    <Zweig>
      <t id="Ta"/>
      <t id="Tb"/>
    </Zweig>
    <Zweig>
      <t id="Tc"/>
    </Zweig>
  </Verzweigung>
  <t id="T2"/>
</Zeitachse>
```

```
<tier type="t">
  <event start="T0" end="Tx">Dauernd</event>
  <event start="Tx" end="T1">unter</event>
  <event start="T1" end="Ta">brichst</event>
  <event start="Ta" end="Tb">Du</event>
  <event start="Tb" end="T2">mich</event>
</tier>
<tier type="a" category="POS">
  <event start="Tx" end="Ta">Verb, 3sg</event>
</tier>
```

# TEI

	T0	T1	T2	T4
A	Dauernd unter	brichst Du mich!		
B		Stimmt doch	gar nicht.	

```
<u who="A" start="T0" end="T2">
  <w>Dauernd</w>
  <w type="V3sg">unter<anchor synch="T1"/>brichst</w>
  <w>Du</w><w>mich</w>
</u>
<u who="B" start="T1" end="T4">
  <w>Stimmt</w><w>doch</w>
  <anchor synch="T2"/>
  <w>gar</w><w>nicht</w>
</u>
```

# Standoff

	T0	T1	T2	T4
A	Dauernd unter	brichst Du mich!		
B		Stimmt doch	gar nicht.	

```
<x id="x1">Dauernd</x>
<x id="x2">unter</x>
<x id="x3">brichst</x>
<x id="x4">Du</x>
<x id="x5">mich</x>
<x id="x6">Stimmt</x>
<x id="x7">doch</x>
<x id="x8">gar</x>
<x id="x9">nicht</x>
```

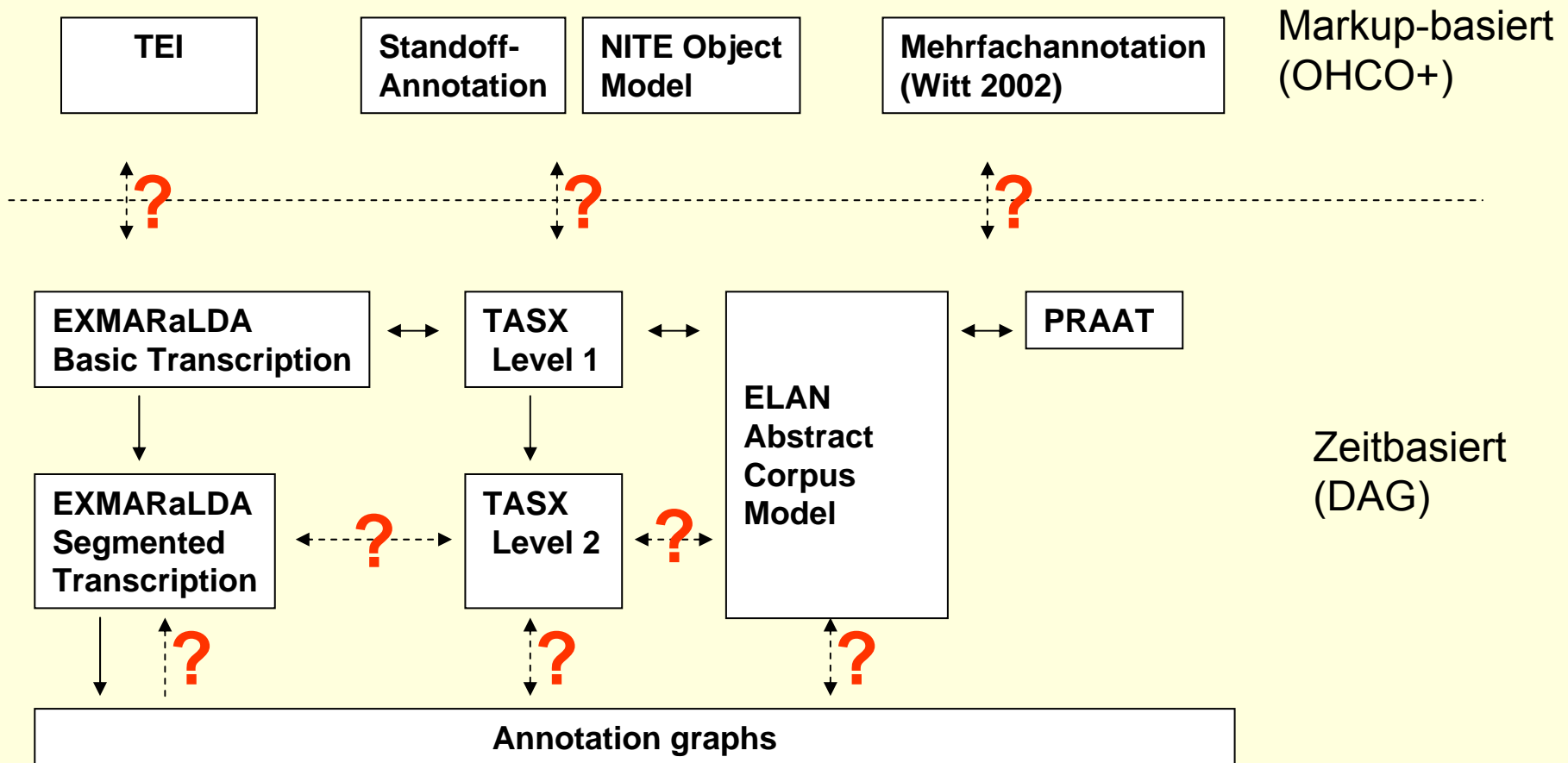
```
<word id="w1" href="x1">
<word id="w2" href="x2...x3">
<word id="w3" href="x4">
<word id="w4" href="x5">
<word id="w5" href="x6">
<word id="w6" href="x7">
<word id="w7" href="x8">
<word id="w8" href="x9">
```

```
<pos href="w2">V3sg</pos>
```

```
<speaker href="x1...x5">A</speaker>
<speaker href="x6...x9">B</speaker>
```

```
<interval>
  <event href="x1...x2"/>
</interval>
<interval>
  <event href="x3...x5"/>
  <event href="x6..x7"/>
</interval>
<interval>
  <event href="x8..x9"/>
</interval>
```

# Datenmodelle, Datenformate, Markuptechniken?



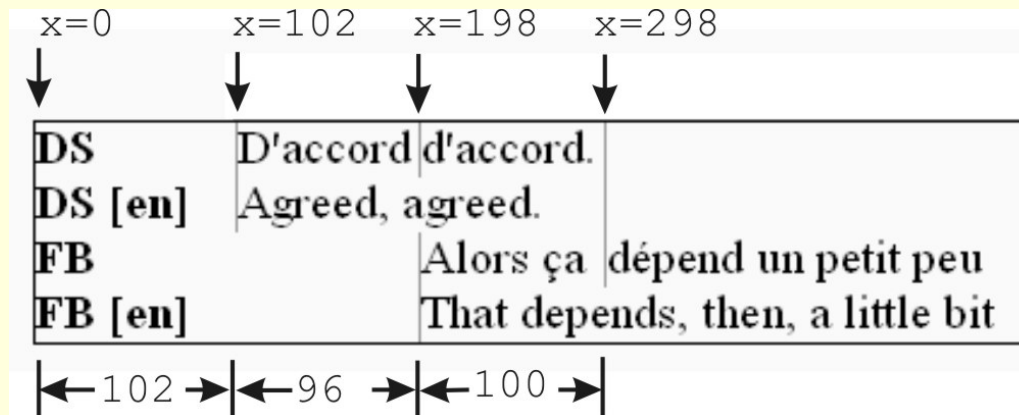
# Software-Bausteine

- Source Code vorhandener Anwendungen
  - TASX
  - ELAN
  - Praat
  - EXMARaLDA
- Bibliotheken (APIs)
  - AGLIB (Anwendung: AGTK)
  - NITE XML Toolkit (Anwendung: "We intend shortly to release three end user GUIs", Jan 05)

# Software-Bausteine für Interlinearen Text

Für die Eingabe:

	0 [1..2]	1	2	3	4	5	6
<b>DS [sup]</b>		<i>faster</i>					
<b>DS [v]</b>	Okay.	D'accord	d'accord.				
<b>DS [en]</b>	Okay.	Agreed, agreed.					
<b>DS [nv]</b>			<i>right hand raised</i>				
<b>FB [v]</b>			Alors ça	dépend ((cough))	un petit peu.		
<b>FB [en]</b>			That depends, then, a little bit				
<b>FB [pho]</b>					[ɛ̃tipø:]		



# Software-Bausteine für Interlinearen Text

Für die Ausgabe:




Diagram showing a single block of interlinear text with a vertical dashed line and a scissors icon on the right side, indicating a cut point.

<b>INT [v]</b>	Do you wanna continue this for the rest of your life, let's say: a musical career?
<b>INT [ger]</b>	Willst Du für den Rest Deines Lebens so weitermachen, mit dieser musikalischen Karriere?
<b>PMC [v]</b>	( ) I don't know; really ((laughs)) ehm, I just wanna be able to do what I want
<b>PMC [ger]</b>	I weiß es nicht wirklich, ich möchte das machen können, wozu ich Lust habe.




Diagram showing two blocks of interlinear text with a vertical dashed line and a scissors icon on the right side, indicating a cut point.

<b>INT [v]</b>	Do you wanna continue this for the rest of your life, let's say: a musical career?
<b>INT [ger]</b>	Willst Du für den Rest Deines Lebens so weitermachen, mit dieser musikalischen Karriere?
<b>PMC [v]</b>	( ) I don't know
<b>PMC [ger]</b>	I weiß es nicht

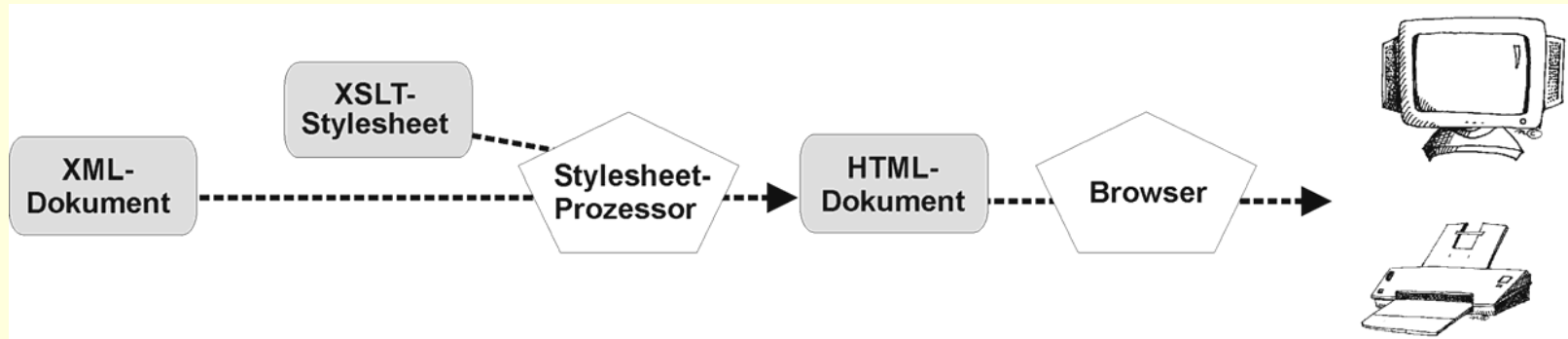
  

<b>INT [v]</b>	really ((laughs)) ehm, I just wanna be able to do what I want to do. And so, you know, at/ at this moment this is what I want to do.
<b>INT [ger]</b>	wirklich, ich möchte das machen können, wozu ich Lust habe. Und im Moment habe ich dazu Lust.
<b>PMC [v]</b>	
<b>PMC [ger]</b>	

USW.

# Software-Bausteine für Interlinearen Text

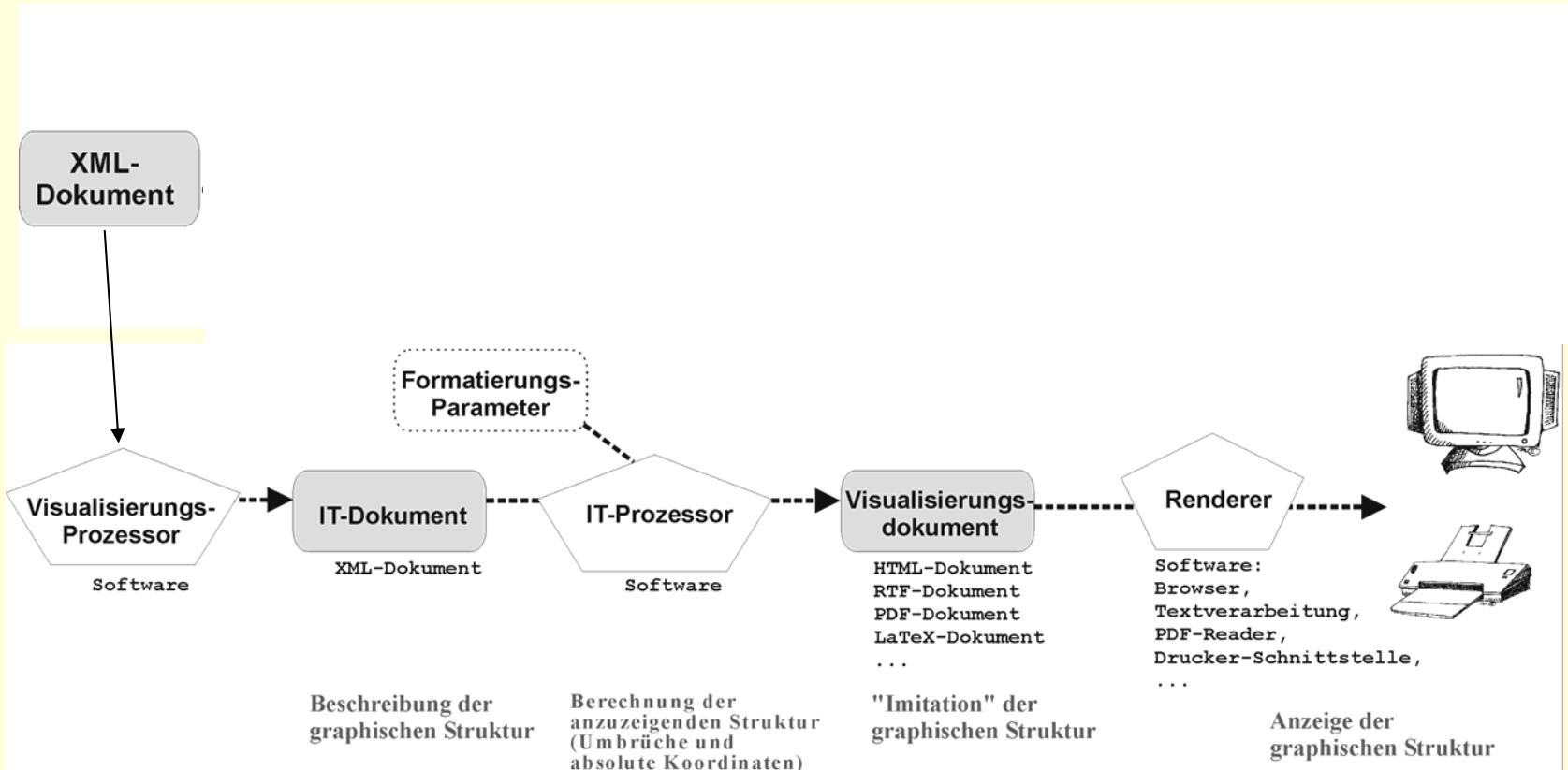
"Linearer" Text



Logische Struktur (Modell)	Visualisierungsverfahren	Graphische Struktur	Rendering	Visualisierung
<pre> &lt;überschrift&gt; Kapitel 1 &lt;/überschrift&gt; &lt;einleitung&gt;   &lt;w&gt;Zunächst&lt;/w&gt;   &lt;w&gt;eine&lt;/w&gt;   &lt;w&gt;Bemerkung&lt;/w&gt; &lt;/einleitung&gt;           </pre>	<ol style="list-style-type: none"> <li>1. &lt;überschrift&gt; → &lt;p&gt;&lt;big&gt;&lt;b&gt;</li> <li>2. &lt;einleitung&gt; → &lt;p&gt;</li> <li>3. &lt;w&gt; →</li> </ol>	<pre> &lt;p&gt;&lt;big&gt;&lt;b&gt;   Kapitel 1 &lt;/b&gt;&lt;/big&gt;&lt;/p&gt; &lt;p&gt;   Zunächst   eine   Bemerkung &lt;/p&gt;           </pre>	Internet Explorer, Safari, Netscape Navigator, Mozilla, etc.	<div style="border: 1px solid black; padding: 5px;"> <p><b>Kapitel 1</b></p> <p>Zunächst eine Bemerkung</p> </div>

# Software-Bausteine für Interlinearen Text

## Interlinearer Text



[1]

	0 [0.]	1 [1.9]	2 [3.2]	3 [4.9]	4 [9.2]
		louder			
X [v]	So it starts out with: A rooster crows. ((1,3s)) ((takes breath 0,5s)) And then you see ehm a maan in maybe his fifties.				
X [nv1]	rHA on rKN, rHA on ISH	rHA up and to the right	rHA stays up	rHA back down rKN, moves rHA on ISH	rHA to the right
X [nv2]			HE nods once		
X [nv3]		emphasizes the crow			approximately fifty
Vid	Shot-1	Shot-2			Shot-3

[2]

	5 [10.4]	6 [11.1]	7 [11.7]	8 [12.4]	9 [13.2]	10 [13.5]	11 [13.7]	12 [13.9]	13 [15.0]	14 [16.0]
X [v]	((takes breath, 0,5s)) And so hee climbsup a tree. And he starts... With a ladder. And he starts picking pearsoff the									
X [nv1]		bHA up, rHA rubs NO	rHA describes a circle	bFA up, in front of BR	rHA goes up	rFA three times up and down in front of BR				
X [nv3]			describes the tree			activity of picking				
Vid			Shot-4			Shot-5				

[3]

	15 [17.0]	16 [18.8]	17 [19.4]	18 [19.9]	19 [20.4]	20 [21.0]	21 [22.4]
X [v]	tree. . . . And he puts the pears into an apron. Okay? ((takes breath, 0,5s)) Aand so then he gathers a bunch.						
X [nv1]	bFA simultaneously, up and down in front of BR		bHA rest in front of BR		rHA rubs NO, rHA down on rKN		bFA move simultaneously in front of CH down to BR
X [nv3]	puts pears into an apron					describes the bunch and the apron	
Y [v]	Okay.						
Vid	Shot-6						Shot-7

# IT-Prozessor in EXMARaLDA

## Eingabe:

IT-Dokument (berechnet. z.B. aus EXMARaLDA Basic-Transcription)  
Formatierungsparameter (z.B. Seitenbreite, "word wrap")

## Berechnung:

absolute Koordinaten  
Umbrüche

## Ausgabe:

HTML-, SVG-, RTF-Dokument, Grafik an Drucker

## Probleme:

Berechnung absoluter Koordinaten: lineares Ungleichungssystem

Berechnung von Umbrüchen:

Typographische Feinheiten, z.B. Ausgleichsverfahren  
("Widow-Orphan-Control" etc.)

Ausgabe: XSL Formatting Objects, PDF

# Weitere Software-Bausteine

- Audio-/Video-Player (für Linguisten!)
  - "Plattformunabhängigkeit"?
  - JMF, QT Java, javax.sound, ...
- Audio-/Video-Player-GUI
  - Navigation in der Aufnahme
  - Waveform Display