

# TalkBank -- Reintegrating the Disciplines

Brian MacWhinney

CMU - Psychology, Modern Languages, LTI,  
SDU - IFKI

# Goals of this talk

- Explain the TalkBank vision
- Explain the core TalkBank principles
- Situate TalkBank within other approaches
- Show in detail what TalkBank has done in 8 specific interest areas

# Historical Integrations

- The Greek Integration
  - Aristotle, Plato
- The Renaissance Integration
  - DaVinci, Descartes, Bacon, Leibniz
- The Modern Integration
  - Systems Theory, AI, Emergentism, Bayes
  - A central role for data-driven mechanisms

# The Challenge

- Grounding integrations on data.
- Multiple topics: classroom, acquisition, code-switching, business, ....
- Multiple methods: linguistics, phonetics, CA, ethology.
- Multiple commitments: linguistics, semiotics, CA
- Weak methods for supporting the integration.

# Principles and Goals

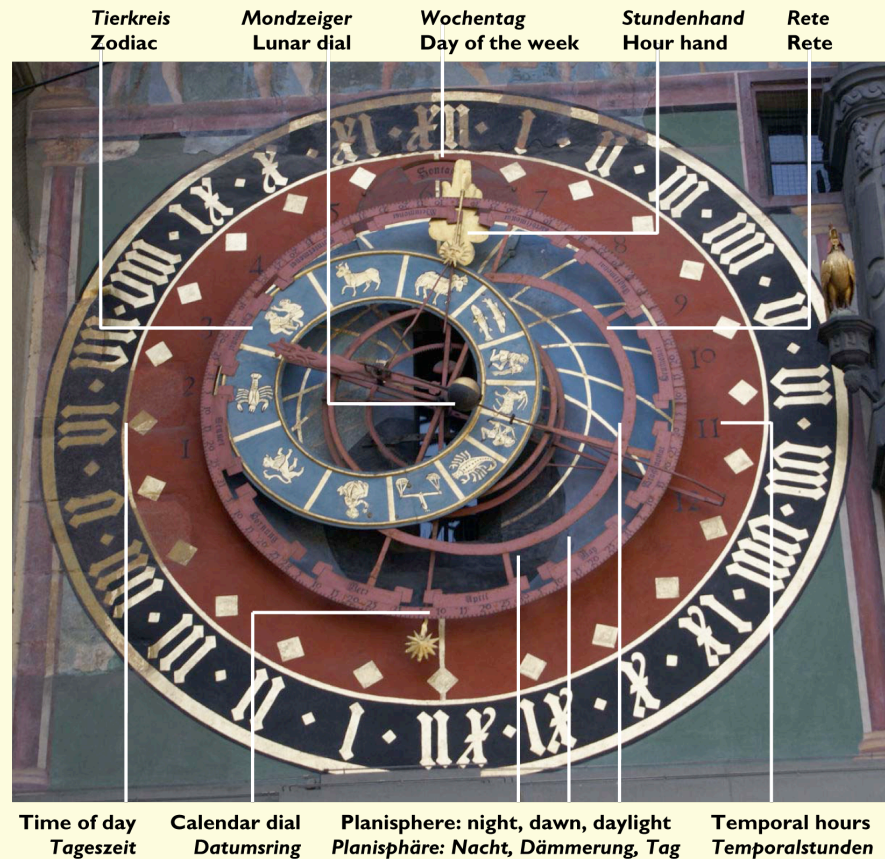
- Data-sharing
- Multimedia
- Open access
- Interoperability
- Integration of disciplines



# The core idea

- Human communication is a single unified process.
- However, patterns in communication are analyzed by 20 different fields.
- The time scales of the processes varies from milliseconds to centuries.
- But all of these processes must have their ultimate effect in The Moment.
- We can capture The Moment on video.

# Meshing of processes



# Available Methods

- Microanalysis - CA, phonetics, ethology
- Microgenetic analysis - Adolph crawling, Icelandic
- Group and treatment comparisons
  - Same consultant, different communities ...
- Error analysis – AphasiaBank, FLLOC ...
- Diffusion analysis - Whiten, Goodwin Professional Vision
- Longitudinal studies - child language, Labov, SCOTUS
- Corpus analysis - Supreme Court, ComNet, LDC
- Modeling - neural nets, dynamic systems, evolutionary models, robotics (iTalk)



# talkbank.org/dreams

- TalkBank + LDC + XXX = ComNet
- Shared Infrastructure for the SBE Sciences
- Other parts of the puzzle
  - CLARIN
  - National Databases (BNC ...)
  - OLAC/IMDI
  - Hundreds of additional corpora
  - Analysis Systems
  - ISO Standards
  - CyberLing

# Current Status

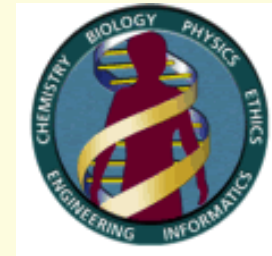
- A web-accessible multimedia database for human communication – browsable (streaming), downloadable, webCLAN corpus analysis
- Interoperability – CLAN, Anvil, CLAPI, ELAN, EXMARaLDA, Praat, etc.
- Integrating tools with group agendas – CHILDES, AphasiaBank, PhonBank, CABank, LIDES, ClassTalk, SamtaleBank, CA/SLA, TBIBank ...
- Transcription standards and interoperability: CHAT2XML Roundtrip

# The ComNet Proposal

- Interoperability
  - Link all the data
  - Link all the standards (AG, LAFS, ISO)
  - Link all the programs
- Open access
- LDC/TalkBank relation: Subset of LDC data in TalkBank format, shared project resources

# Large databases

- Human Genome Project
  - 3 billion base pairs
- Sloan Digital Sky Survey
  - 100 million stars
- Alzheimers Neuroimaging
  - 800 patients over 3 years
- fMRI Data Center
- The Human Connectome



**ADNI** Alzheimer's Disease Neuroimaging Initiative

The fMRI Data Center  
**fMRIDC**

# ComNet within Science

- The Dream Database
- Regional data centers
- Fractal sampling methods
- Integrating ComNet with the Census
- Linking SCOTUS to the Law
- Legacy Databases
- Digital objects

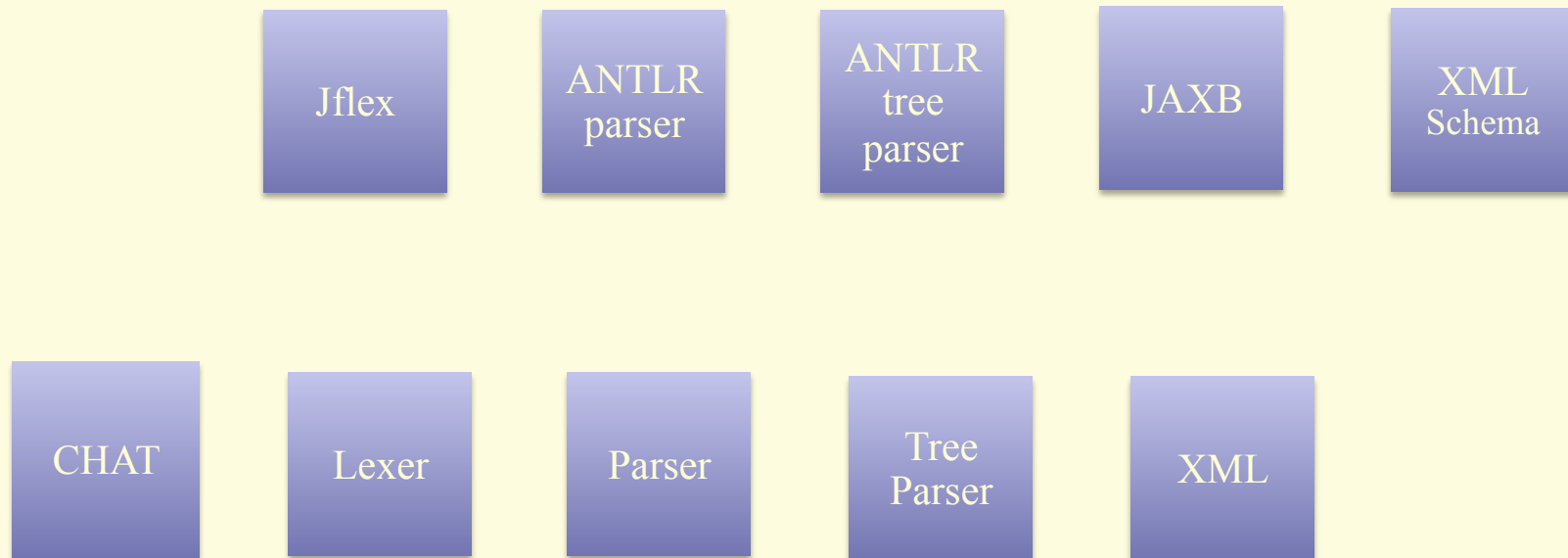
# Availability

- <http://chilides.psy.cmu.edu>
- <http://talkbank.org>
- programs, manuals, fonts, morphologies, CA conventions, video production guides, XML Schema, roundtrip, converters to other programs
- Interactive or downloadable
  - Browser
  - WebData
  - WebCLAN

# CHAT - A common format

- CHAT
- XML
- Reformatting from other formats to XML
- Round trip involves
  - CHAT2XML
  - XML2CHAT, comparison
  - Java with ANTLR, JFLEX, JAXB
  - Automatic checking and validation

# RoundTrip





# CLAN Tools

- Transcribing
- Editing
- Counts -- `FREQ`, `KWAL` and 23 others
- Morphosyntax: `MOR`, `GRASP`
- Phonology: `Praat`, `PHON`
- Interoperability -- `ELAN`, `Praat`, `Anvil`, `EXMARaLDA`, `CLAPI`, `PHON`

# 8 target areas

1. Child Language
2. PhonBank
3. SLA - LIDES
4. Aphasia
5. CA - CLAPI
6. Gesture/Multimodality
7. Classroom
8. Professional Discourse - Legal, Medical

# 1 - Child Language

- Child Language Data Exchange System
- Founded 1984 in Concord MA
- Director: Brian MacWhinney [macw@cmu.edu](mailto:macw@cmu.edu)
- Programmers: Leonid Spektor, Franklin Chen, John Kowalski
- 3200 Members
- 140 corpora
- Over 3200 published articles

# CHILDES and TalkBank

	CHILDES	TalkBank
Age	24 years	8 years
Words	44 million	8 + 55 million
Media	2 TB	.5 TB
Languages	33	18
Publications	3200+	89
Users	3200	600

# Why study child language?

- Special Gift and Universals
- Emergentism – Processes and Mechanisms
- Typological Effects
- Language Disorders
- Socialization, Literacy
- Language Maintenance

# Universals

- Are there basic patterns to babbling? - Davis
- Are early word orders universal? - Pinker
- Does UG give children a universal set of functional categories?
- Is the vocabulary spurt universal? - Bates

We need LOTS of data

# Differences

- Do children have individual styles?
  - Gestalt vs. Analytic - Peters
  - Enactive (1S) vs. Depictive (3S) - Dressler
- Is child learning shaped by parental input? - Snow, Bohannon, MacWhinney
- Do cultures vary in what they teach? - Leonard, Kokusai, Shanghai, Mexico

Again, we need LOTS of data.

# Speech Act Coding

```
@Begin
@Languages: en
@Participants: MOT Mother, CHI David Target_Child
@ID: en|rollins|MOT||||Mother||
@ID: en|rollins|CHI|1;8.||||Target_Child||
@Activities: book
*MOT: ahhah: look we can read books Tim .
%spa: $DHA:YY $DHA:RP
*MOT: it's a look and see <book> [>] .
%spa: $DHA:ST
*MOT: <ahhah> [>] we open it up and there are a set of eyes
%spa: $DJF:ST $DHA:ST
*MOT: <the bear has a baby> [>] bottle .
%spa: $DHA:ST
*MOT: yes # David has baby <bottles> [>] .
%spa: $DRP:ST
*MOT: <oh> [>] .
%spa: $DHA:MK
*MOT: <there's a mirror> [>] .
%spa: $DJF:ST
*MOT: can David see <David> [>] .
%spa: $DHA:RQ
*CHI: 0 .
%act: CHI look-s at rattle in hand then puts rattle in mouth
```

http://xml.talkbank.org:8888/talkbank/roll...

[0:13:19 - 0:19:19]

\*MOT: <there's a mirror>

%spa: \$DJF:ST

■ ■

User Comment 1: A good coding sequence with high correlation to the research community

User Comment 2: I refer the reader to Keck et. al where you will see that this coding scheme is highly disputed.






# Morphosyntax

- POS tagging for 12 languages. With correct transcription, accuracy is at 98%
  - MOR generates tags
  - POST disambiguates
  - POSTMORTEM examines residual issues
- GRASP uses output of MOR to add dependency structure with 38 relations. English, Japanese, Hebrew, Spanish

# 2. PhonBank

PhonMedia      EnglishDemoCorpus : Project Manager

Media   Segment



–IPA Actuals transcribed : Session Editor

hday parties?)  
 θ,deɪ 'pɑ:ti:z]

s ɪ ŋ æ t 'b ʌ ɹ θ d eɪ p ɑ ɹ t i: z

s ɪ ŋ æ t 'b ʌ ɹ θ d eɪ p ɑ ɹ t i: z

Orthography    [I sing] [happy birthday] [to you] [horsie]

IPA Target      ['aɪ 'sɪŋ] ['hæpi: 'bʌɪθ,deɪ] ['tu: 'ju:] [hɔ:si:]

Target Syllables    a ɪ s ɪ ŋ h æ p i: b ʌ ɹ θ d eɪ t u: j u: h ɔ: ɹ s i:

IPA Actual        ['sɪ:] ['hæpi: 'bʌ:,teɪ] ['tu: 'ju:] ['hɔ:si:]

Actual Syllables    s ɪ h æ p i: b ʌ: t eɪ t u: j u: h ɔ: s i:

Alignment        a ɪ s ɪ ŋ h æ p i: b ʌ ɹ θ d eɪ t u: j u: h ɔ: ɹ s i:  
                          s ɪ h æ p i: b ʌ: t eɪ t u: j u: h ɔ: s i:

Segment         000:02.843 to 000:07.843

# Corpora

- English – Davis
- French – Lyon, Leonard, Kern
- German – Lleo, Stuttgart
- Japanese – Ota
- Romanian – Kern
- Dutch – Levelt/Fikkert, Zink
- Québécois – Rose/Goad

## 3. SLA

- Childhood bilingualism -- 14 corpora
- CUHK Bilingual corpus
  - Evidence for early transfer, language mixing
  - Problems for modular theories
- ESF MPI Gastarbeiter corpora
- FLLOC for French
- Barcelona Spanish corpus Díaz-Rodríguez
- CASL

# Alicia at 2;7

/Volumes/BriPod/video/Child/AliciaCan/ac020713.cha

Movie - Sound

\*SIS: 好呢個咩嚟架? •  
%mor: adv|hou2=very sfp|ne1=sfp cl|go3=cl  
wh|me1=what sfp|lei4=sfp sfp|gaa3=sfp ?

\*SIS: 呢個咩嚟架? •  
%mor: det|ni1=this cl|go3=cl wh|me1=what  
sfp|lei4=sfp sfp|gaa3=sfp ?

\*SIS: [- en] toy . •  
%mor: n|toy .

\*CHI: [- en] toy . •  
%mor: n|toy .

\*SIS: 一個 . •  
%mor: num|jat1=one cl|go3=cl .

\*CHI: xx . •  
%mor: unk|xx .

\*SIS: 姐姐 嚟 lu3 . •  
%mor: n|ze4=sister sfp|lei4=sfp sfp|lu3=sfp .

\*CHI: 姐姐 嚟 lu3 . •  
%mor: n|ze4=sister sfp|lei4=sfp sfp|lu3=sfp .

\*SIS: 哎地 風箏 嚟 囉 . •  
%mor: co|ai1aa3 n|fung1zang1=kite sfp|lei4=sfp sfp|bo3=sf



7 2002  
15:15

Save

# Adult Bilingualism

- Bangor
- BlumSnow
- Eppler
- Hatzidaki
- Langman
- Etc.

# 4. Aphasia

## System

[Ground rules](#)

[Membership list](#)

Subscribing to the [Mailing List](#)

## Protocol - Testing Materials

[Instructions](#)

[Troubleshooting](#)

[Repetition Test](#)

[Demographics Coding Sheet](#)

[Verb Naming Test](#)

## Materials

[IRB Guidelines](#)

The AphasiaBank [NIH Proposal](#)

[Video Recording Guidelines](#)

[December Meeting Notes](#)

## Phonology and Fonts

The [Phon](#) program and database

Unicode and IPA for [Mac](#)

Unicode and IPA for [Windows](#)

## Database

[Transcript Database](#)

[Media Database](#)

[Playback without downloading](#)

## Protocol - Additional Materials

[Participant Inclusion Criteria](#)

[Protocol List](#)

[Protocol Description](#)

[Test Results Excel Sheet](#)

[Demographics Excel Sheet](#)

## Programs and Manuals

The [CLAN](#) Program

[CHAT Transcription](#)

[Training Materials](#)

[CLAN Programs](#)

## Transcription

[Error Coding](#)

[Foot Pedal for CLAN](#)

[Coding Cheat Sheet](#)

# Corpus sustainability

- CHILDES is embedded in child language
- AphasiaBank will be fundamental to aphasia research
- Can this be done for other fields?
- Does this block integration?



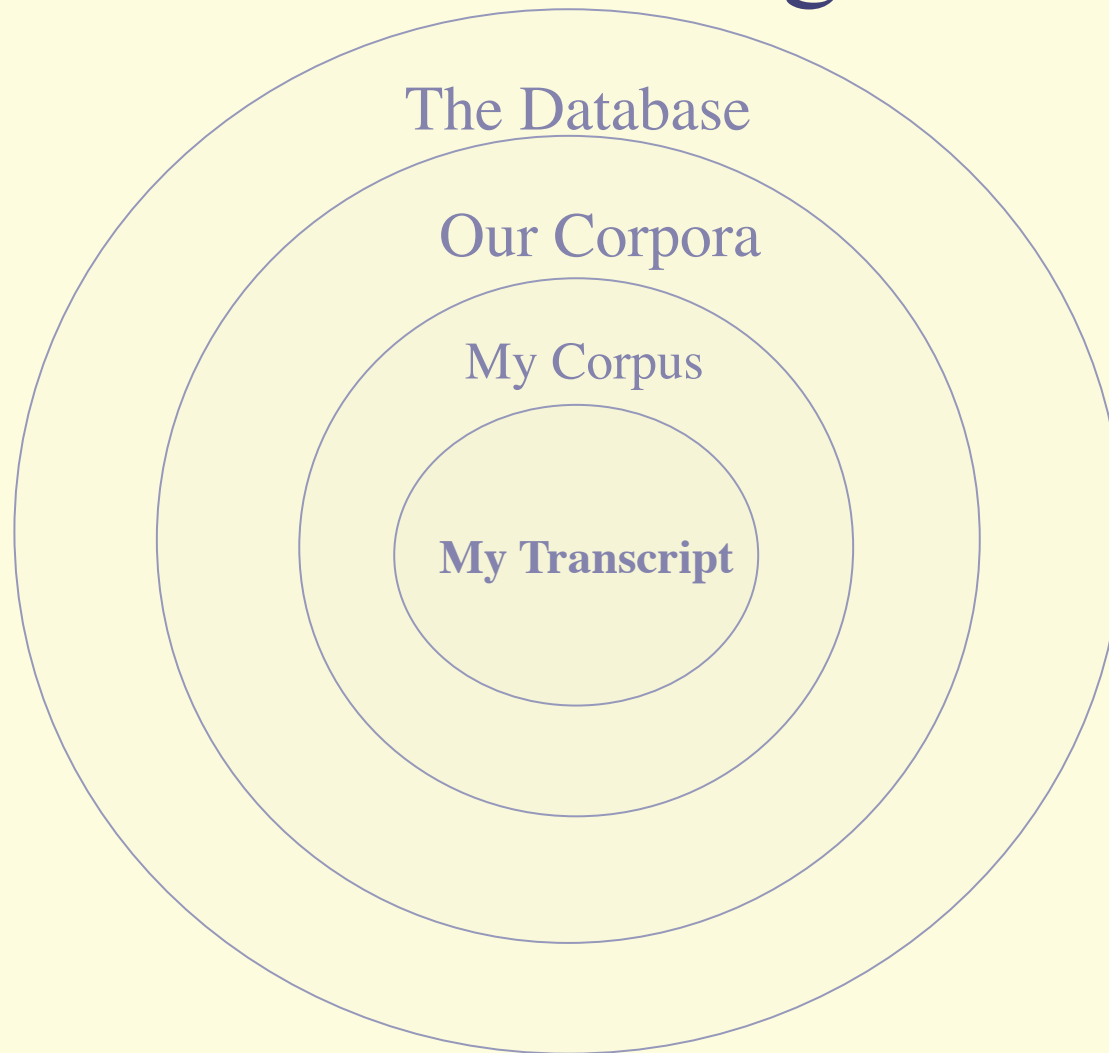
## 5. CA

- Johannes Wagner - MOVIN
- Jefferson Corpora
- SLA/CA
- Medical
- CLAPI
- CallFriend
- Santa Barbara
- DK/CLARIN

# Challenges

- Digging deeper into the dynamics of conversation
- Avoiding premature generalizations
- Maintaining an evidential collection
- Promoting pattern searches
- Downward integration - prosody, phonetics, multimedia
- Upward integration - syntax, learning, discourse

# Circles of Usage



# Transcripts linked to media

The screenshot displays a software interface with two main windows. The left window, titled "Movie - Sound", shows a video of three people in a meeting. The right window, titled "Coyote:demo:MyTheory.ca", displays a transcript of the video content. Below the video player are playback controls and a "Repeat" button.

**Movie - Sound**

**Coyote:demo:MyTheory.ca**

```
1 @Begin
2 @Transcriber: Tim. Koschmann. Last revision 8.1.2000 Johannes Wagner
3 @Participants: Be Betty, No Norman, Co Coach, Mar Maria, May, Jen Jenny,
4 Lill, ? unidentified Person, Ps Pauses
5 @Dependent: ges
6 @Filename: MyTheory.ca. Moviefile MyTheory.mov
7 @Time: 6 minutes
8 @Contents: fragment of tutor-group disocssion
9 @Comment: numbering is by TCUs and pauses, not lines
10
11 Be: See what it said in here (.) in- my theory (hhh) •
12 (0.4)
13 ?: khu- [(.hhh)
14 Be: [about this amnesic (.) dysnomic aphasia, •
15 (0.3)
16 Be: u:hm (it) says the cause of lesion is usually deep in
17 temporal lobe just like Kathy was saying ↑presumably
18 interrupting connections of sensory speech areas with the
CLAN [E][CA] 1
```

1425

0 3200

Repeat b0 msec

MYTHEORY.MOV

# CA marks in Unicode

<u>Char</u>	<u>Function</u>	<u>F1 +</u>	<u>Unicode</u>
↑	shift to high pitch	up arrow	2191
↓	shift to low pitch	down arrow	2193
ˆ	rising, but not to top	?	00BF
˘	falling, but not to bottom	;	037E
·	inhalation	.	2219
≈	≈latching	=	2248
⌈	top begin overlap	[	2308
⌋	top end overlap	]	2309
⌌	bottom begin overlap	shift [	230A
⌍	bottom end overlap	shift ]	230B
↗	↗ faster ↗	right arrow	2197
↘	↘ slower ↘	left arrow	2198
*	* creaky *	*	2605
√	√louder√	/	221A
°	° softer °	zero	00B0
—	— low pitch —	d	2581
—	— high pitch —	h	2594
£	£ smile voice £	l	00A3
¢	pulse of laughter or breath	c	00A2

# Additional facilities

- Underlining
- Overlap alignment
- CAFont fixed-width font
- Heritage level coding
- CLAPI searching

# Searchable Features

Cutoffs	+/.
Overlaps	┌ ┐ └ ┑
Fillers	um, em
Pauses, pause length	(.) (6.2) or #6_2
Repeats, retraces	[/] [//]
Prosodic	↑ ↓ ↘ ↗
Latching	≈ +,
Paralinguistic	&=
Others	— —

## 6. Gesture



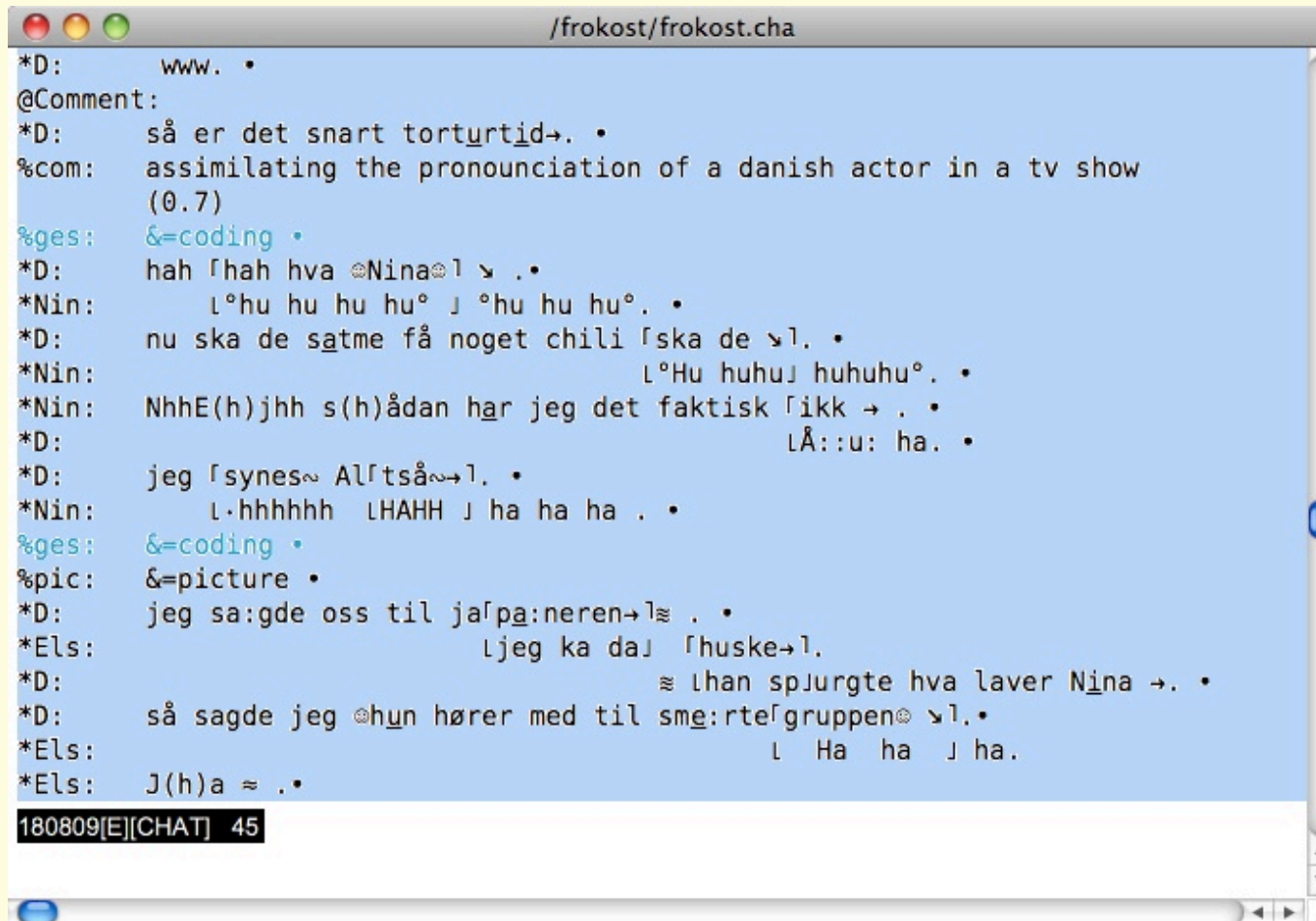
5/22/04

TalkBank for Conversation

40



# Overall transcript



```

/frokost/frokost.cha
*D:      www. •
@Comment:
*D:      så er det snart torturtid→. •
%com:    assimilating the pronunciation of a danish actor in a tv show
(0.7)
%ges:    &=coding •
*D:      hah [hah hva @Nina@l v] . •
*Nin:    [°hu hu hu hu° ] °hu hu hu°. •
*D:      nu ska de satme få noget chili [ska de v]l. •
*Nin:    [°Hu huhu] huhuhu°. •
*Nin:    NhhE(h)jhh s(h)ådan har jeg det faktisk [ikk →] . •
*D:      [Å::u: ha. •
*D:      jeg [synes~ Al[tså~→]l. •
*Nin:    [h h h h h h] [HAHH ] ha ha ha . •
%ges:    &=coding •
%pic:    &=picture •
*D:      jeg sa:gde oss til ja[pa:neren→]≈ . •
*Els:    [jeg ka da] [huske→]l.
*D:      ≈ lhan spJurgte hva laver Nina →. •
*D:      så sagde jeg @hun hører med til sme:rte[gruppen@ v]l. •
*Els:    [ Ha ha ] ha.
*Els:    J(h)a ≈ . •
180809[E][CHAT] 45

```

# Converting

> chat2elan mytheory.cha

Mon Jul 9 02:56:29 2007

chat2elan (28-Jun-2007) is conducting analyses on:

ALL speaker tiers

and those speakers' ALL dependent tiers

and ALL header tiers

\*\*\*\*\*

From file <mytheory.cha> to file <mytheory.eaf>

Number of tiers found: 177

Number of missing bullets: 0

Please look at file "mytheory.err.cex"

Done with file <mytheory.eaf>

- Three parts
- Each part has components
- Each part linked
- Each part displayed

```

/frokost/2smertegruppen.cut
Sequence: 3 part •

#Part 1 •
Body part 1      torso
Class            orientation
Action          leaning forward
Direction       Nina

Body part 2      elbow
Class            orientation
Direction       Nina
Action          place, support
Object          table

Body part 3      bottom
Class            beat - jaPANeren
Action          sit down
Object          bench

*D:      jeg sa: gde oss til ja[pa:neren→l]≈ •
%1:      ≈torso_lean_forward-----]•
*Els:                                Ljeg ka da J

#Part 2 •

Body part      torso
Class          orientation
180809[E][TEXT] * 1

```

# In CHAT and CLAN

The screenshot displays a software interface with two main windows. The left window, titled "Movie - Sound", contains a video player showing a group of people in a meeting. Below the video player are playback controls, including a volume slider set to 1425, a progress bar from 0 to 3200, and a "Repeat" button set to "b0" msec. The filename "MYTHEORY.MOV" is visible at the bottom of this window. The right window, titled "Coyote:demo:MyTheory.ca", displays a transcript of a discussion. The transcript is numbered 1 through 18 and includes metadata such as "@Begin", "@Transcriber: Tim. Koschmann. Last revision 8.1.2000 Johannes Wagner", "@Participants: Be Betty, No Norman, Co Coach, Mar Maria, May, Jen Jenny, Lill, ? unidentified Person, Ps Pauses", "@Dependent: ges", "@Filename: MyTheory.ca. Moviefile MyTheory.mov", "@Time: 6 minutes", "@Contents: fragment of tutor-group disocssion", and "@Comment: numbering is by TCUs and pauses, not lines". The transcript text includes: "11 Be: See what it said in here (.) in- my theory (hhh) • (0.4)", "12 ? : khu- [(.)hhh)", "14 Be: [about this amnesic (.) dysnomic aphasia, • (0.3)", "16 Be: u:hm (it) says the cause of lesion is usually deep in temporal lobe just like Kathy was saying ↑presumably interrupting connections of sensory speech areas with the". At the bottom of the chat window, the text "CLAN [E][CA] 1" is visible. A red question mark icon is located at the bottom center of the interface.

# In ELAN

The screenshot displays the ELAN software interface for a file named "Elan - mytheory.eaf". The interface includes a menu bar (File, Edit, Annotation, Tier, Type, Search, View, Options, Window, Help) and a toolbar with playback and editing controls. A video window on the left shows a classroom scene with a teacher and students. To the right of the video are sliders for Volume (0-100) and Rate (0-200). Below the video is a selection bar showing the current selection range: "00:00:40.530" to "00:00:41.515 985".

The main transcription area shows a timeline from 00:00:41.000 to 00:00:49.000. The transcription is organized into tiers:

- \*BET (Background)
- \*UNK (Unknown)
- \*NOR (Non-overlapping)
- \*COA (Contextual)
- %gpx@NOR (Gloss)
- \*MAR (Main transcription)
- %gpx@MAR (Gloss)
- %gpx@COA (Contextual)

The transcription text is as follows:

Time	*MAR	%gpx@MAR	%gpx@COA
00:00:41.000 - 00:00:42.000	if you lift up +/-	brings R hand in	
00:00:42.000 - 00:00:43.000	that little temporal lobe	lifts R hand above head	
00:00:43.000 - 00:00:44.000	[insid		
00:00:44.000 - 00:00:45.000	#0_		
00:00:45.000 - 00:00:46.000	Middle top?		
00:00:46.000 - 00:00:47.000			
00:00:47.000 - 00:00:48.000			
00:00:48.000 - 00:00:49.000	0.	María poin	

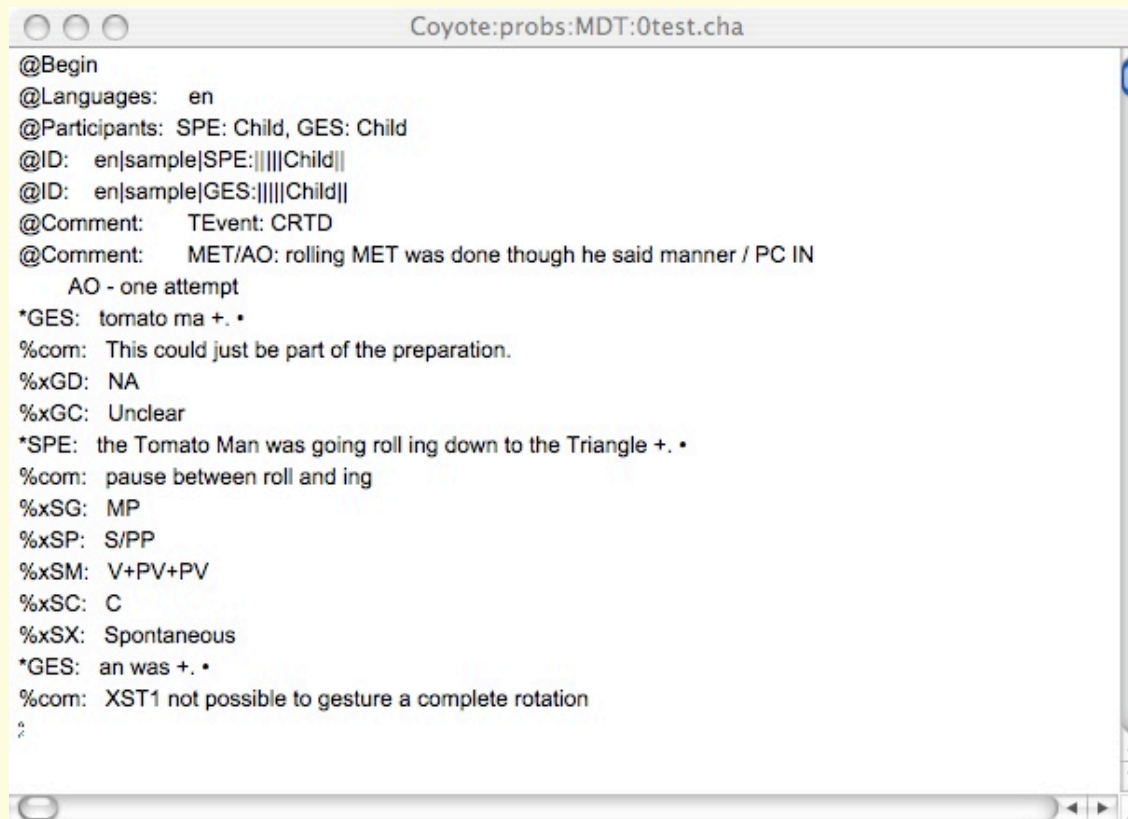
Additional annotations include: "You can you can point to it on" (00:00:43.000 - 00:00:44.000) and "Points with R hand from seat t" (00:00:46.000 - 00:00:47.000).

# ELAN2CHAT

- It looks the same as the original. This shows that we have a “roundtrip”.
- But if you edit in ELAN, some features may no longer be “good CHAT”. You can run CHECK to fix these.
- Or you can ignore CHECK, if the data will not be in the database.

# The Kita/MPI corpus

MDT2CHAT, CHAT2ELAN, ELAN2CHAT



```
Coyote:probs:MDT:0test.cha
@Begin
@Languages:  en
@Participants:  SPE: Child, GES: Child
@ID:  en|sample|SPE:||||Child|
@ID:  en|sample|GES:||||Child|
@Comment:  TEvent: CRTD
@Comment:  MET/AO: rolling MET was done though he said manner / PC IN
      AO - one attempt
*GES:  tomato ma +. •
%com:  This could just be part of the preparation.
%xGD:  NA
%xGC:  Unclear
*SPE:  the Tomato Man was going roll ing down to the Triangle +. •
%com:  pause between roll and ing
%xSG:  MP
%xSP:  S/PP
%xSM:  V+PV+PV
%xSC:  C
%xSX:  Spontaneous
*GES:  an was +. •
%com:  XST1 not possible to gesture a complete rotation
^
```

# In ELAN

The screenshot displays the ELAN software interface for a video file named "Elan - 0test.eaf". The interface includes a menu bar (File, Edit, Annotation, Tier, Type, Search, View, Options, Window, Help), a video player window showing a child, and a transcription timeline. The timeline is divided into several tiers:

- \*GES**: Contains phonetic segments: "tomato ma +.", "an w", "was +.", "rolling do +.", "own to the tria +."
- %com@GES**: Contains annotations: "This could just be XST XST2 p", "XST1 ?", "XST2 ? Hard ges"
- %xGD@GES**: Contains "NA" annotations.
- %xGC@GES**: Contains "Unclear", "M", "P" annotations.
- \*SPE**: Contains the transcription: "the Tomato Man was going roll ing down to the Triangle +."
- %com@SPE**: Contains "pause between roll and ing"

The timeline also shows a selection range from 00:00:03.068 to 00:00:03.569. The video player shows a volume slider at 100 and a rate slider at 100. The current time is 00:00:04.299.



# In EXMARaLDA

EXMARaLDA Partitur-Editor 1.3.3p1 [untitled.xml]

V+PV+PV

	1 [1.8]	2 [2.7]	3 [3.0]	4	5 [3.9]
<b>*GES</b>	1 +.	an was +.	was +.		rolling do +.
<b>%com@GES</b>	ast be part of the preparation.	XST1 not possible to gesture a complete rotation	XST2 pause and change in hand shape not significant enough to split gesture		XST1 ?
<b>%xGD@GES</b>		NA	NA		NA
<b>%xGC@GES</b>		M	P		M
<b>*SPE</b>	the Tomato Man was going roll ing down to the Triangle +.				
<b>%com@SPE</b>	pause between roll and ing				
<b>%xSG@SPE</b>	MP				
<b>%xSP@SPE</b>	S/PP				
<b>%xSM@SPE</b>	V+PV+PV				
<b>%xSC@SPE</b>	C				
<b>%xSX@SPE</b>	Spontaneous				

Done.

# Hong Kong Sign Language

Elan - sample.eaf

File Edit Annotation Tier Type Search View Options Window Help

Grid Text Subtitles Controls

Volume: 100

Rate: 100

00:00:34.826 Selection: 00:00:34.826 - 00:00:36.984 2158

Selection Mode Loop Mode

	000	00:00:29.000	00:00:30.000	00:00:31.000	00:00:32.000	00:00:33.000	00:00:34.000	00:00:35.000	00:00:36.000	00:00:37.000
*BRE	DER	SISTER	IX_3 ?	IX_1 WHO	IX_1 ?	IX_3 ?	SAME .	IX_1 ELDER_SISTER	IX_3 ELDER_SI	
%xgl1@BRE	ELDE	IX_3 ?	I	WHO	IX_1 ?	IX_3 ?	SAME.	IX_ >	IX_3	SAME.
%mor@BRE	_3 n ELDER_SISTER n:		n:pro IX_1	n:pro IX_1 ?	n:pro IX_3 ?	adj SAME .	n:pro IX_1 n ELDER_SISTER n:pro IX_			
%xgl2@BRE							ELD   ELDE			

# 7. Classroom Discourse

- TIMMS - six countries
- PBL - Koschmann, LeBaron
- Gravity - TERC
- Science Museum
  - Atmospheric light diffusion - Rahm
  - Electricity generation - Crowley
- Dresden - SLA English, French, Czech
- Grimshaw - Oral Defense
- Greeno - Garden Plot, numerical series

# Classroom - continued

- Numerical displays - Sfard, McClain, Cobb
- Lehrer - Carmen Curtis and quilt patterns
- Lectures -- MacWhinney gesture analysis
- Science/Math classes and discovery
- Home/School -- Hall, Snow
- ClassTalk for Teacher Training

# Tutorial Interactions

- Circle - Physics
- Frederiksen - statistics
- Graesser - statistics
- DISPEL - collaborative problem solving

# Sample Analyses

- James Greeno, Brian MacWhinney, and Carla van der Sande
- Learning as the construction of mental models that explain device representations.
- Humans represent (explain) devices through
  - Perspectival embodiment
  - Spatial imagery

# Gabriel's Model



# Dad's Model



5/22/04

TalkBank for Conversation

56



# Gravity and Pprimms



5/22/04

TalkBank for Conversation

57

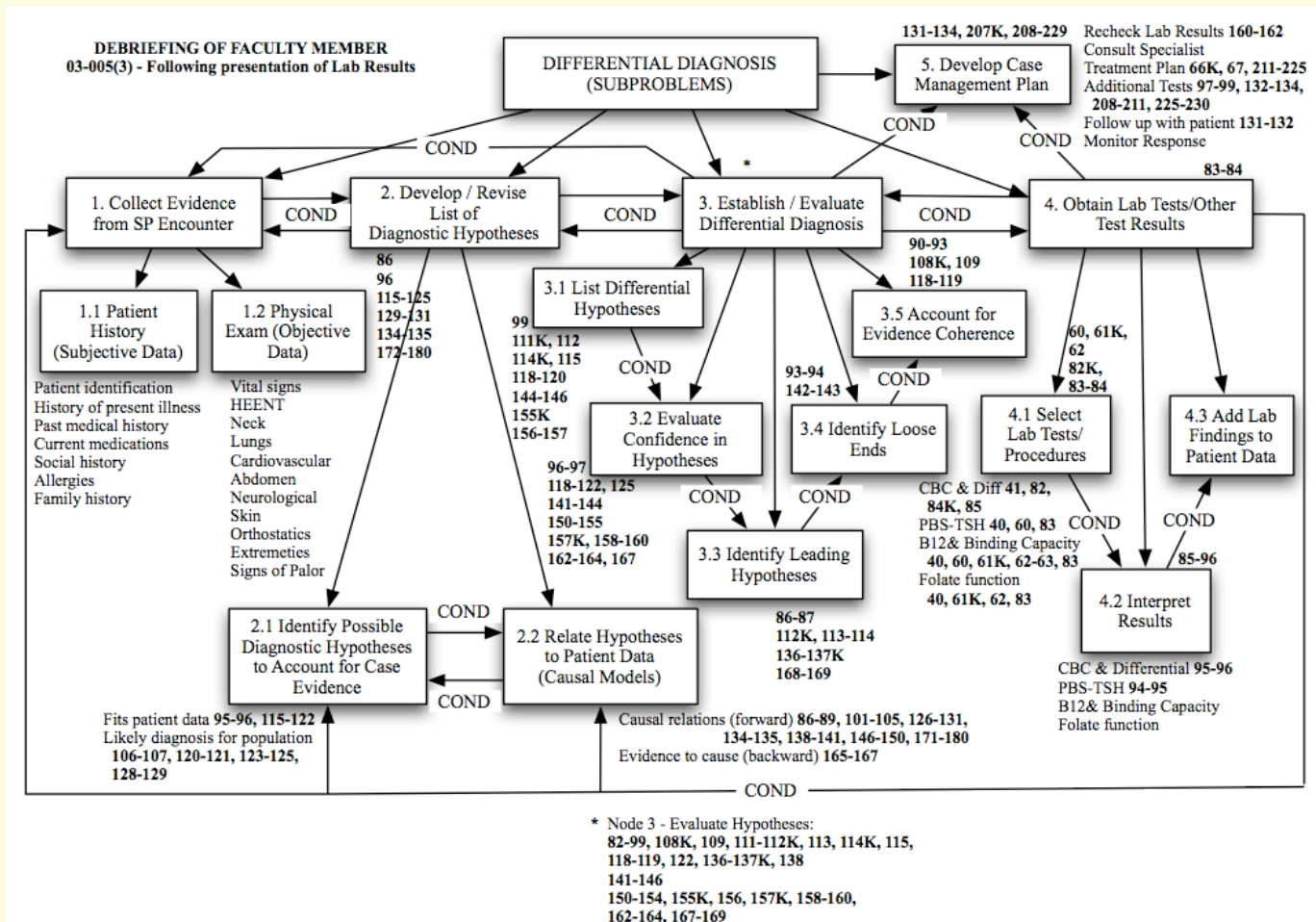
# Commitment to full database

- We must construct complete propositional tree analysis.
- Coding must be reliable.
- Model must apply across all TalkBank datasets, across subjects.

# 8. Professional Discourse

- Supreme Court and Decisions
  - The SCOTUS work group
- Medical Examinations and Competency
  - The Competency Project
- Collaborative Commentary

# Polycythemia - Frederiksen



# Comment Tagging, Filtering

- Automatic: author, date, media begin-end
- Author self-characterized metadata (role, faction, position, credentials)
- Commentary type (refutation, defense, elaboration, analogy, statistics, case law, gesture-speech match)
- Filters: only teacher, only from colleagues, etc.

# Naked Video

- Terabytes of video
  - Speechome, Classroom, Resident Care
- No transcripts
- Occasional sign posts
- Sparse speech recognition
- Automatic video analysis

# Conclusions

- We may achieve a new integration
- But we still need to provide the technical basis for data-sharing, interoperability, and collaborative commentary
- After that, the major barrier is a full commitment to data-sharing
- And patience to integrate across seven time scales.

# Postscript: Automatic Analyses

