

# Development of sign language acquisition corpora

Diane Lillo-Martin\* & Deborah Chen Pichler<sup>°</sup>

\*University of Connecticut and Haskins Laboratories;

<sup>°</sup>Gallaudet University

# Longitudinal spontaneous corpora

## Benefits

- Allows observation of long-term language development, in-depth detail
- Amenable to testing a wide range of hypotheses within a variety of theoretical frameworks
  - theory neutral
  - additional coding can always be added to reflect new insights, theoretical assumptions, etc.
- Complements experimental/cross-sectional study nicely

# Longitudinal spontaneous corpora

## Drawbacks

- MacWhinney's (2001) three-headed monster of corpus transcription:
  - Lack of standard format + rapid proliferation of alternative formats
  - Indeterminancy
    - Difficult to determine what was really said/signed
  - Tedium
    - Highly labor-intensive, continually subject to revision and expansion

# Longitudinal spontaneous corpora

## Desiderata (cf. MacWhinney 2001)

- multiple children, across time
- various familiar interlocutors
- captures representative sample of child's normal language
- strikes a balance between transcription thoroughness and time/tedium
- allows easy access to points of interest, for analysis or additional coding
- facilitates collaborative commentary

# Uconn-Gallaudet Corpus



- Longitudinal spontaneous production data
  - various adult interlocutors interact with children
- L1 ASL signers:
  - Monolingual Deaf
  - Bilingual coda
- Ongoing transcription over last 10 years

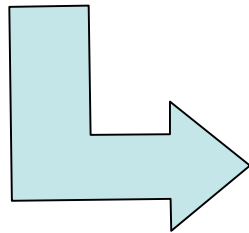
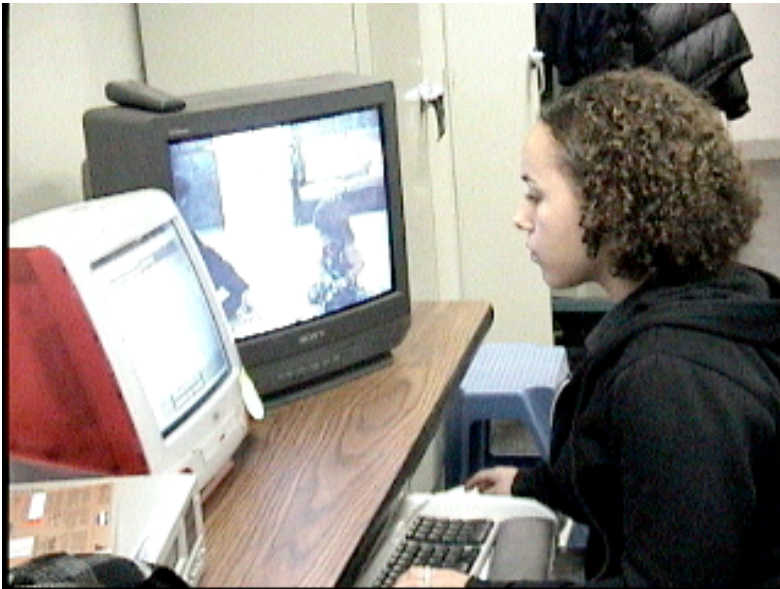
# ASL acquisition corpora

Uconn and Gallaudet combined sessions

	Child	Age Range	#Sessions	#Hours (approx)
<b>ASL</b>				
<b>D/D</b>	Abby	1;05 – 3;04	79	75
	Jill	1;07 – 3;07	77	79
	Ned	1;05 – 4;02	44	40
	Sal	1;07 – 2;10	18	16
<b>D/H</b>	Cal	6;10 – 10;01	115	50
	Mei	6;07 – 10;0	111	50
<b>H/D</b>	Ben	1;04 – [4;04]	[100]	[80]
	Tom	1;04 – [4;04]	[100]	[80]
	Pete	1;07 – [4;07]	[100]	[80]

# Previous incarnation of database

## FileMaker Pro + Autolog



JIL65.fmp.done

Stop Transcribing

Coder	Date
① DSM	
②	
③ SCG	10/20/2000

ENTER  
GO TO

REWIND STOP PLAY FORWARD

Records Previous Next New

Utterance Time on videotape 00:02:30:17

Signer JIL

Manual Gloss <MYSELF> <POUND-WITH-HAMMER> <MYSELF μPICK-UPμ>

HamNoSys Gloss

Eye Gaze < > camera towards floor camera

Action BRO is using hammer, JIL stretches both hands out towards

Comment FAT has his tools out, including hammer

Facial Expression	1st Instance	2nd Instance	3rd Instance	4th Instance
[ ] Head				
f f Brow				
≤ ≥ Eyes				
{ } Nose				
μ μ Mouth	teeth clench			
/ \ Shoulder				

# Previous incarnation of database

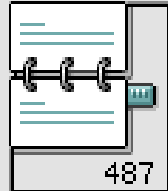
## FileMaker Pro + Autolog

- FMP database designed for utterance-by-utterance coding
- VCR controlled within FMP via Autolog
- Multiple layouts specifically for:
  - coding of sign
  - coding of context
  - coding of nonmanual information
- Search and printing features



Stop Transcribing

Coding



Records:  
719

Sorted

	Coder	Date
①	LAL	08/24/19
②	JMT	06/05/20
③	SCG	07/07/20

ENTER  
GO TO

REWIND STOP PLAY FORWARD

R < < < < < || > > > > > H

Records Previous Next New

Utterance

Time on videotape 00:12:50:03

Signer

Manual Gloss

HamNoSys Gloss

Eye Gaze < >

Action

Comment

Facial Expression

	1st instance	2nd instance	3rd instance	4th instance
[ ] Head	shake			
/ / Brow				
< > Eyes				
{ } Nose				
μ μ Mouth	shhh			
/ \ Shoulder				

# Current incarnation of database

## ELAN

- Quicktime (H.264 codec) movies in ELAN
- Tiers for:
  - adult and child sign
  - adult and child speech (for bilinguals)
  - comments, feedback, analysis
- Elected to continue using English glosses
- Building of database
  - coding done at both Uconn and Gallaudet
  - transcripts available to small group of Uconn and Gallaudet students/faculty
  - “enriched” transcripts returned tagged to database



00:04:44.755

Grid Text Subtitles Controls

Child Sign

FATHER? · WHAT[B-o] · DADDY · IX(KNEE) · FATHER · ME · ME · FALL · ME · IX(KNEE) · FATHER · IX(ME) · NO? · FALL · NO · LIFT · ME · IX(KNEE) · LIFT · WHAT[B-o] · LATER · IX(OC) · NO? · WITH · MOTHER · PAINT · IX(OC) · YES? · WRITE · HALF · IX(OC) · GREEN · FORK? · SPOON? SOUP? · HALF? · IX(OC) · EAT · SOUP? · SMALL? · IX(OC) · BATH · IX(OC) · PAINT · WAIT-A-MIN? O-K? · CAN · XXX · PAINT · IX(OC) · PAINT · IX(BACK OF ROOM?) · PHONE? · IX(OC) · WAIT-A-MIN · WELL · COLOR · XXX · WAIT · WATER · WAIT · WATER? · WAIT · COLOR · WATER · IX(OC) · LET-ME-SEE · IX(BOOK) · IX(PAINTBOOK) · ME · IX(BOOK) · ME · IX(BOOK) · WAIT-A- · GIRL · (GES) · (GES) · NO · IX(PAINT) · XXX · **TENT? CAMPING?** · WHITE · HOUSE>? · TENT? · IX(OC) · COOL? · HAVE · IX(OC) · BIG · IX(OC) · DAD · XXX · FINISH · ME · EAT · DON'T-KNOW · XXX · XXX · FINISH · ME · FINISH · NO? · BIG · EAT · BIG · FINISH · IX(PAINTING BOOK) · XXX · IX(PAINT PICTURE) · IX(PAINT PICTURE BOOK) · ME · ME · IX(PAINT BOOK) · YYY · IX(PAINT BOOK) · XXX · IX(PAINT BOOK) · GIRL? · IX(PAINT BOOK) · IX(PAINT BOOK) · IX(PAINT BOOK) · XXX · IX(PAINT BOOK) · LEAVE-IT · WET · IX(PAINTING) · IX(PAINTING) · XXX · HER · MOM · IX(PAINTING) · IX(PAINTING) · HER · MOM · WAIT-A-MIN · WAIT-A-MIN · RED · SAME · MINE? · XXX · XXX · (GES) · (GES) · DRINK · A · B · C · D · E · NEXT? · F · G · C · A · IX(OC) · A · C · C · FORGET · IX(OC) · C · C

Selection: 00:00:00.000 - 00:00:00.000 0



	00:04:40.000	00:04:41.000	00:04:42.000	00:04:43.000	00:04:44.000	00:04:45.000			
<b>Child Utterance/Tr</b> [362]						A tent.			
Child Sign [647]						TENT? CAMPING?			
Gesture-C [15]									
<b>Adult 1 Utterance/T</b> [0]									
Adult 1 Sign [1800]		CAMPING		CAMPING	QUESTION	CAMPING	IX(SAL)		WH
Gesture-A1 [18]									
Comment [78]									
Repetition-A1 [54]						++			
Repetition-C [52]									

# Benefits of current system

- Linkage to video allows less detailed basic transcript
- Encourages collaborative commentary-  
new tiers easy to add or hide
- Merge function facilitates collaborative transcription of the same video files
- Various output options facilitate analysis

# Sample analysis in Excel

	A	B	C	D	I	J	K	N	O	
1	Time Code	Full utterance	Interpretation	Predicate	Error	Pred Type	Mood	Signs / Utt	Morphs / Utt	C M
2	00:00:13:00	<IX-ctr FATHER IX-ctr GIVE-ME>	Daddy, give me (that)!	GIVE-ME	no	person	imperative	4	6	
3	00:00:19:00	<μIX(CRR)μ BOOK [WANT ME IX(bag)]>	I want (the/that) book.	WANT	no	plain+loc	assertion	5	6	
4	00:00:52:00	GET	He will get (the chair)	GET	no	person+loc	irrealis	1	3	
5	00:00:54:00	IX(fat) GET	He will get (the chair)	GET	no	person+loc	irrealis	2	4	
6	00:00:55:00	<CHAIR? IX(chair) GET, WAIT-A-MIN>	(He's) getting the chair, wait a minute	GET	no	plain	irrealis	4	4	
7	00:00:55:00	<CHAIR? IX(chair) GET, WAIT-A-MIN>	(He's) getting the chair, wait a minute	WAIT-A-MIN	no	emblem	imperative			re
8	00:00:59:00	<WAIT>	You wait...	WAIT	no	plain	imperative	1	1	
9	00:01:00:00	<CL:5-PUT-THERE> <CHAIR WAIT-A-MIN>	He will put the chair right there.	CL:5-PUT-THERE	no	CL-SASS	irrealis	3	4	
10	00:01:00:00	<CL:5-PUT-THERE> <CHAIR WAIT-A-MIN>	He will put the chair right there.	WAIT-A-MIN		emblem	imperative			re
11	00:01:01:00	<COME [WAIT-A-MIN]>	He's coming; wait a minute	COME	no	person+loc	irrealis	2	4	
12	00:01:01:00	<COME [WAIT-A-MIN]>	He's coming; wait a minute	WAIT-A-MIN	no	emblem	imperative			re
13	00:01:05:00	WAIT-A-MIN	Just wait a minute..	WAIT-A-MIN	no	emblem	imperative	1	1	
14	00:01:07:26	<THERE(chair)>, OVER-THERE BRING, CL:claw5-PUT-DOWN>	There it is! He is bringing it and putting it down over there.	BRING	no	person+loc	assertion	4	7	
15	00:01:07:26	<THERE(chair)>, OVER-THERE BRING, CL:claw5-PUT-DOWN>	There it is! He is bringing it and putting it down over there.	CL:claw5-PUT-DOWN	yes	CL-SASS	assertion			re

# Remaining practical challenges

- Optimal methods for video compression and archiving
- Management of wider data sharing and collaborative commentary
- Consent and levels of confidentiality for wider sharing (especially for child data)
- ELAN-specific issues
  - modification of tier dependencies
  - vertical rather than horizontal view
  - remove default format for media files

# Reference

MacWhinney, Brian (2001) From CHILDES to TalkBank, In Research on Child Language Acquisition, Almgren, M. Barreña, A. Ezeizaberrena, M., Idiazabal, I., and MacWhinney, B. (Eds.), Cascadia: Somerville, MA. pp. 17-34

# Acknowledgments

- National Institutes of Health, National Institute on Deafness and Other Communication Disorders, Grant #DC-00183
- National Science Foundation, Division of Behavioral and Cognitive Sciences, Grant #BCS-0078788
- University of Connecticut Research Foundation
- Gallaudet Research Institute
- Both groups wish to thank the signing children and their families who have graciously allowed us to film them over the last fifteen years, as well as the numerous research assistants who have been crucially involved in collecting and particularly in transcribing the data.