Sign Language Acquisition: Annotation, Archiving and Sharing



(SLAAASh project) – Status Report

Diane Lillo-Martin^{1,2}, Lee Prunier², Julie Hochgesang^{2,3}, Deborah Chen Pichler^{2,3} ¹University of Connecticut, ²Haskins Laboratories, ³Gallaudet University



Introduction

GOAL

• Prepare corpus of previously-collected data on ASL acquisition to share with other researchers

WHY?

Shared data will permit more researchers to conduct studies of ASL acquisition and use

Dataset

UConn CLESS Child ASL Data (Lillo-Martin & Chen Pichler 2008)

- Spontaneous production data from 4 Deaf children of Deaf parents, ages 1;04-4;01
- Interlocutors: Deaf parents; hearing, signing experimenters

Child	# Sessions	Age begin	Age end	Time observed (hrs:mins)	Est. # gloss tokens	Est. # child utterances
ABY	79	1;04.22	3;04.07	73:43	130,000	16,600
JIL	83	1;07.03	3;07.09	79:16	119,000	17,800
NED	44	1;05.28	4;01.28	40:00	60,000	9,000
SAL	18	1;07.18	2;10.01	17:11	23,000	3,900
Total	224			210:10	332,000	47,300

What kind of corpus?

"a collection of (1) machine-readable, (2) authentic texts which is (3) sampled to be (4) representative of a particular language or language variety" (McEnery, Xiao & Tono 2006: 5)

Existing shared ASL data include



Reconsenting

(Chen Pichler et al. 2015, 2016)

• Need to request consent for data sharing

Tools (with Onno Crasborn)

<u>ASL SignBank</u> – Lexicon of ID glosses (cf. Johnston 2001) – currently ~2200 signs

ASL signbank										
	t (Diane Lillo-Martin)	arch gloss	Search translation	Sign search						
	File Edit Annotation Tier Type Search	/iew Options Window Help	4.9.4 - SAL_001_aes_MLUc0_170316.eaf							
ELAN bridge – Connection between ELAN		Child free transle		ts Recognizers Metadata Controls						
 and SignBank SignBank fields visible in ELAN Pop-up window with gloss options Edit SignBank from within ELAN Automatic updating of SignBank glosses 	0: 19: 25: 6 0: 19: 25: 6 10: 19: 24. 125	67 ft its a pumpkin. 68 ft is a dog. 69 ft is a cat. 70 ft is a bear. 71 There. 72 Hey, 1am looking. 73 There. 74 Move it there. 75 There. 71 Can't do this. 77 I can't do this. 78 This.well. 79 Hey, there, there 80 Hey. 81 It is red. 82 There. 83 Hey. 84 There. 85 This is dirty. 86 There. X8 There. X8 There. X8 There. X8 There. X9 Hey. 84 There. 85 JS → X8 There. X8 There.	g.	00:17:29.175 00:17:29.175 00:17:41.665 00:00:01.16 00:17:54.665 00:17:56.165 00:00:01.50 00:19:21.205 00:19:21.205 00:19:21.205 00:19:22.425 00:19:21.205 00:19:22.425 00:19:21.205 00:19:21.415 00:00:02.6175 00:19:38.660 00:20:00.910 00:20:01.830 00:20:01.830 00:20:01.830 00:20:01.830 00:20:01.830 00:20:01.830 00:20:01.830 00:20:01.830 00:20:01.830 00:20:01.830 00:20:01.830 00:20:01.830 00:20:01.835 00:20:21.415 00:00:01.241 00:20:32.460 00:20:31.455 00:00:02.51 00:22:10.865 00:22:11.935 00:20:31.455 00:00:02.63 00:22:10.865 00:22:1.235 00:25:12.235 00:26:21.235 00:26:21.235 00:26:21.235 00:26:50.340 00:26:50.340 00:26:50.340 00:26:50.340 00:27:07.145 00:00:01.261 00:27:05.885 00:27:07.145 00:00:01.261						
Automatic updating of Signballk glosses	00:19:20.000 00:1	<u>a nome i i i i i i i i i i i i i i i i i i i</u>	00:19:25.000 00:19:26.000 00:19:27.000 00:19:28.000 00:1	11111						
 Automatic POS tagging 	Child free translation Hey, I am looking. [86] Child ASL right hand [99]	IX		Move it there.						
• Other ELAN improvements for ease of use	Child ASL RH append Child ASL RH append Child ASL right hand pho (11) Child ASL left hand Child ASL Left hand Child ASL LH append (13) Child ASL Left hand pho (13) Child ASL left hand pho Child ASL left hand pho (14) C	(floor)		(toy)						

from all participants

- Seek ethically-sound, community-supported practices for decision-making
- Focus groups for input

from various types of stakeholders

- Protection of individual rights primary concern; research potential important but secondary
- Give participants option for different levels of sharing
- Policies for special cases: segments to be edited out; faces blurred

 \bigcirc



Progress Report

- Continuing to revise annotation \bullet conventions for internal/ external consistency, partnered with ASL-LEX (Caselli et al., 2016)
- Currently converting old annotation files to new system
- Completing annotation of previously unfinished files
- Populating ASL SignBank with new signs and search aids

 \bigcirc

(55) — Child MLUm (55)			2	4
Child MLU Comments	i() not counted	IX only		
[36]				

• Conducting basic descriptive analyses (MLU, IPSyn)

- Completed two focus groups to establish reconsenting process guidelines
- Completed "bridge" linking ASL SignBank to ELAN allowing for direct, controlled annotation of our glosses

Future Plans

- Release each data set as it is prepared
- Share tools open source
- Share video only when annotations not available

 \bigcirc

Literature cited

Caselli, N., Sevcikova, Z., Cohen-Goldberg, A., Emmorey, K. (2016). ASL-Lex: A Lexical Database for ASL. *Behavior Research Methods*.

 \bigcirc

Websites of some existing

 \bigcirc

SL corpora

 \bigcirc

http://hsldb.georgetown.edu/ http://blackaslproject.gallaudet.edu/ BlackASLProject/Welcome.html

Acknowledgments

 \bigcirc



Research reported here was supported in part by the National Institute on Deafness and other Communication Disorders of the National Institutes of Health under Award Number R01DC013578. The content is solely the

Chen Pichler, D, Hochgesang, J & Lillo-Martin D

(2015) Digging Workshop

Chen Picher, D, Hochgesang, J, Simons, D & Lillo-Martin, D (2015) TISLR

Chen Pichler et al. (2016) LREC

Johnston, T (2001) Sign Language & Linguistics 4 Lillo-Martin & Chen Pichler (2008) LREC

http://www.auslan.org.au/about/ corpus/

http://www.bslcorpusproject.org/

https://signbank.csc.fi/

http://www.sign-lang.unihamburg.de/dgs-korpus/

http://www.ru.nl/corpusngten/

responsibility of the authors and does not necessarily represent the official

views of the National Institutes of Health.

 \bigcirc







 \bigcirc