Sign Language Acquisition: Annotation, Archiving and Sharing (SLAASh project) – Status Report

Diane Lillo-Martin[1,2], Lee Prunier[2], Julie Hochgesang[2,3], Deborah Chen Pichler[2,3]
[1]University of Connecticut, [2]Haskins Laboratories, [3]Gallaudet University

Introduction

GOAL

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

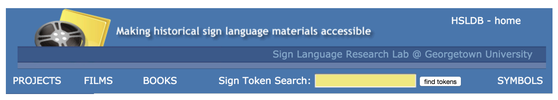
WHY?

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

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 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

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

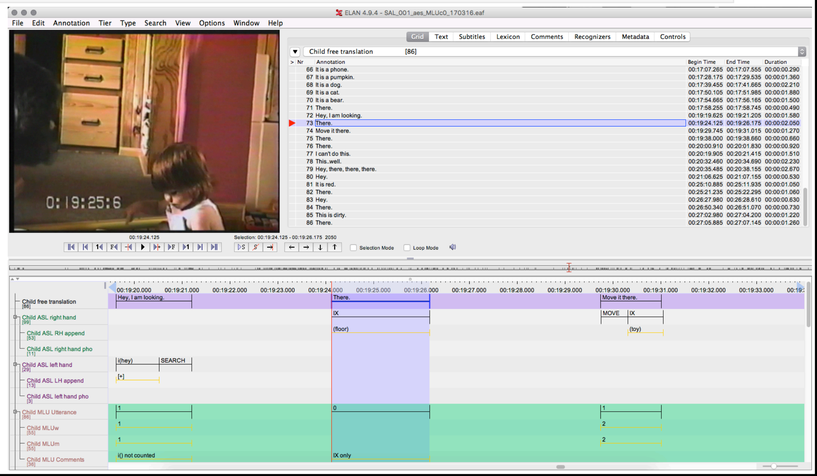
Child	# Sessions	Age begin	Age end	Time observed (hrs:mins)	Est. # gloss tokens	Est. # child utterances
ABY	79	1;04.22	3;04.07	73:43	130,000	16,600
JIL	83	1;07.03	3;07.09	79:16	119,000	17,800
NED	44	1;05.28	4;01.28	40:00	60,000	9,000
SAL	18	1;07.18	2;10.01	17:11	23,000	3,900
Total	224			210:10	332,000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

Tools (with Onno Crasborn)

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

ELAN bridge – Connection between ELAN and SignBank

- SignBank fields visible in ELAN
- Pop-up window with gloss options
- Edit SignBank from within ELAN
- Automatic updating of SignBank glosses
- Automatic POS tagging
- Other ELAN improvements for ease of use



Progress Report

- Continuing to revise annotation 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
- Conducting basic descriptive analyses (MLU, IPSyn)
- Completed two focus groups to establish re-consenting 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

Literature cited

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

Websites of some existing SL corpora

http://hsldb.georgetown.edu/
http://blackaslproject.gallaudet.edu/BlackASLProject/Welcome.html
http://www.auslan.org.au/about/corpus/
http://www.bslcorpusproject.org/
https://signbank.csc.fi/
http://www.sign-lang.uni-hamburg.de/dgs-korpus/
http://www.ru.nl/corpusngten/

Acknowledgments

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 responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.