BIBIBI Project ASL Annotation Conventions

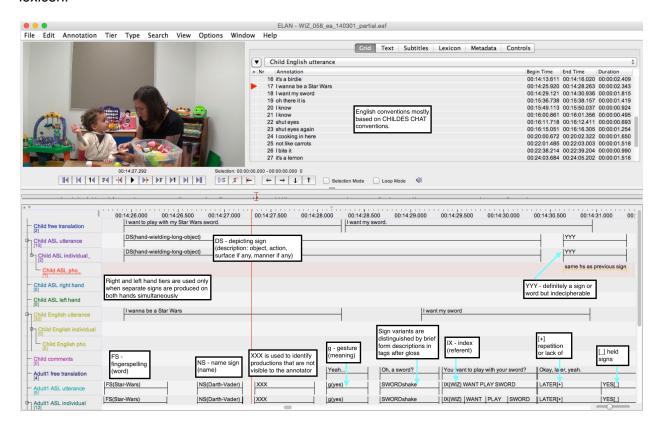**Deborah Chen Pichler[1], Julie Hochgesang[1], Diane Lillo-Martin[2,3]**

[1]*Gallaudet University, *[2]*University of Connecticut, *[3]*Haskins Laboratories*[1]

Introduction

These conventions have been developed by a team of researchers at Gallaudet University and the University of Connecticut (GUC) (see Chen Pichler et al. 2010). The primary purpose is for the annotation of longitudinal spontaneous production data from Deaf children of Deaf parents and from bimodal bilingual hearing children of Deaf parents. Data collection and annotation is on-going.

Annotation conventions take into consideration our analysis goals, and our attempt to use a format as consistent as possible with both common sign language annotation symbols and those used in CHILDES (MacWhinney 2000). The initial annotation passes focus on glossing of signs and words as well as translations. We are working toward use of an ID Gloss lexicon.

Annotator responsibilities

We aim to ask annotators to do as little analysis as possible. Our intention is to include basic information in the sign tiers, with additional details about use of space, non-manuals, etc. left to subsequent analysis passes. This principle guides our decisions. Annotators should use ID glosses and use the translation tier for further information about their interpretation of each utterance.

[1] deborah.pichler@gallaudet.edu; julie.hochgesang@gallaudet.edu; diane.lillo-martin@uconn.edu

Tier Hierarchy

Our tier hierarchy starts with the ASL Utterance, with automatic tokenization for the ASL Individual tier based on spaces between annotations. RH/LH are used only when necessary (simultaneous constructions). This provides information from annotators about utterance groupings, and saves time compared to annotation of all two-handed signs twice.

Capitalization

Capitalization systematically represents a conventionalized sign (e.g., SIGN) or type of sign (e.g., DS (for depicting sign) or FS (for fingerspelling)). Lower case signifies that additional information is presented. For conventionalized signs, the lower case portion that follows the capitalized gloss indicates something about its form when distinguishing between variants (SOONnose, SOONchin). For information enclosed in parentheses following codes, aspects of the meaning expressed by the sign is presented in lower case. From a human-readability perspective, this allows the user to perceive patterns in the data just from scanning. Non-sign communicative acts (when annotated) use lower-case (e.g., show(toy)).

Partly/non-lexical material

Following our principles, upper-case codes identify the type of sign, with lower-case information added, or supplementary information in additional tiers. For example, the annotation for a depicting sign indicates its category (DS), and the additional information conveys a rough approximation to meaning: DS(vehicle-goes-down-street). Pointing signs are annotated using IX(referent). Further analysis takes place independently on separate tiers.

References

Chen Pichler, D., Hochgesang, J.A., Lillo-Martin, D., & Quadros, R. (2010). Conventions for sign and speech transcription in child bimodal bilingual corpora. *Languages, Interaction and Acquisition* 1, 11-40.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Table 1: Comparison of BIBIBI, BSL and NGT Conventions

	Description of our approach	Our example	BSL example	NGT example
Basic gloss	Lexical signs are written using ID glosses (no spaces)	SIGN	SIGN	SIGN-A
Two-handed signs	RH and LH tiers are only used when the hands are producing different signs	SIGN DIFF-SIGN	SIGN SIGN	SIGN SIGN
Buoys	Use of RH and LH tiers	SIGN DIFF-SIGN	ID gloss	ID gloss
Lexical variants	Information about form in ID glosses used to distinguish between variants	DOGslap DOGsnap	DOG DOG02	DOG-A DOG-B
Repetition	Signs with atypical # of repetitions marked [+]; held signs marked [_]	GLOSS[+] GLOSS[_]	SIGN SIGN	SIGN SIGN
Compounds	No special annotation	SIGN	SIGN^SIGN	SIGN-SIGN
Manual negative incorporation	Individual ID glosses	DON'T-KNOW	KNOW-NOT	KNOW-NOT
Directional verbs	Verbs are written using ID glosses with no annotation for directionality	VERB	(none)	ASK:1
Plurality		(none)	(none)	GLOSS.PL
Numbers	Numbers written in words	ONE	ONE	1
Number sequences	Each term has its own ID gloss	TWENTY-ONE	ONE^TWO	12
Number incorporation	Each term has its own ID gloss	FIVE-WEEKS	HOUR-FOUR02	HOUR-4
Ordinal numbers	No special annotation	FIRST	FIRST	1-ORD
Sign names	NS followed by the name in parentheses	NS(name)	SN:FIRST	*FIRST-LAST
Fingerspelling	FS followed by the perceived word in parentheses	FS(perceived-word)	FS:WORD FS:WORD(WRD)	#WORD
Pointing signs	IX followed by the referent (or location) to which the point is directed. IX(self) is used for pointing to self. POSS(referent) and SELF(referent) are used for pointing signs with the B and A-dot handshapes respectively.	IX(referent) POSS(referent) SELF(referent)	PT:LOC PT:DET PT:PRO1SG PT:LOC/PT:PRO3SG PT:LBUOY	PT PT:B PT:W PT:1
Classifier/depicting signs	DS followed by a description of the depiction	DS(object-description-surface)	DSEW(2)-MOVE DSEP(1)-PIVOT DSEW(2)-A T	MOVE+2 ('cat walks to and fro') PIVOT+1 ('cat's legs move around') AT+2 ('bird is here')
Shape constructions	Same as above	(see above)	DSS(CYL)	SHAPE+cylinder 'drain pipe'
Type-like classifier/depicting signs	Same as above	(see above)	DSEW(1-VERT)-MOVE:HUMAN DSEW(FLA T-LA TERAL)-AT:VEHICLE	MOVE+1 MOVE +flat
Gestures	Lower-case g followed by the name of the gesture. For interjections, i(interjection) is used.	g(gesture) i(interjection)	G:HOW-STUPID-OF-ME	% HEY

	Description of our approach	Our example	BSL example	NGT example
Palm up	ID gloss PU	PU	G:WELL	
Manual constructed action	As action (e.g., &=poses)	&=action	G:CA:HOLD-HANDS-UP-IN-FRIGHT	%
Doubt as to whether sign	XXX: the annotator cannot identify any part of the sign; YYY: the annotator can describe the form (on the pho tier)	XXX or YYY or GLOSS[?]	INDECIPHERABLE	±
Correct gloss chosen?	GLOSS[?] indicates uncertainty; GLOSS[=?ALT-GLOSS] indicates an alternative option	YYY or GLOSS[?]	?GLOSS or GLOSS1/GLOSS2	?GLOSS
Annotator doesn't know sign; needs to be double-checked	No special annotation	(none)	ADD-TO-SIGNBANK(UNKNOWN)	??
None of us knows this sign	YYY as gloss with description of the form on the pho tier	YYY		
Proposal for new gloss	No special annotation	(none)	ADD-TO-SIGNBANK(GLOSS)	$GLOSS
New gloss needed, no proposal yet	No special annotation	(none)	-	$
Invisible, but likely sign		XXX	?GLOSS	!GLOSS
Invisible, unclear or doubtful which sign it is	See above alternatives	XXX or YYY or GLOSS[?]	Options given separated by / , or INDECIPHERABLE	!
False start, but the sign is recognised as GLOSS	Interruption GLOSS/ Self-interruption GLOSS// Retracing without correction GLOSS[/] Retracing with correction GLOSS[//] Retracing with reformulation GLOSS[///]	GLOSS/ GLOSS// GLOSS[/] GLOSS[//] GLOSS[///]	GLOSS(FALSE-START)	~GLOSS
False start, not clear what the sign was going to be	Use uncertainty symbols with false start symbols	GLOSS[?]// or YYY// etc	INDECIPHERABLE(FALSE-START)	~
Annotation filenaming convention	Pseudonym uses 3 upper case letters File# uses 3 digits Uppercase E,A indicate annotated and checked (English, ASL) Lowercase e,a indicate annotated but not yet checked Date is latest annotation revision	Pseudonym_File#_ (EA)_YYMMDD.eaf	Region + Participant_number + Gender + Age + Ethnicity + Deaf/ hearing_family + Task (e.g. LN01M25WDL)	Corpus + Number (e.g. CNGT0001) Metadata tiers will be created to facilitate searching for annotations + metadata
Video filenaming convention	see above	Pseudonym_File#. mov	Region + Participant_number(s) + task (e.g. L1l; CF5+6c)	Corpus + Number + Participant_number + View (e.g. CNGT0001_S003_b)
Video file viewing convention	One annotation file per session	(none)	One annotation file per signer	One annotation file per signer pair
Parsing	String is divided by utterance; individual glosses by tokenization (plus optional adjustment)	Annotations are true to frames.	Small gaps (2 frames) between annotations	Annotations are true to frames.