Stockholm University

Swedish Sign Language Corpus

Digging into Signs Workshop: Developing Annotation Standards for Sign Language Corpora, London, March 30-31, 2015

Lars Wallin & Johanna Mesch

Stockholm University

Swedish Sign Language Corpus, SSLC

- Dataset
  - 24 hours of video data (300 media files)
  - 42 native and near-native signers (age 20-82)
  - Semi-spontaneous and elicited dialogues
- Approx. 23% (69 media files, 4:48 hours)
  - Glosses (approx. 42,000)
  - Swedish translations (approx. 62,000 words)
- Approx. 17% (52 media files, 4:46 hours)
  - Swedish translations (approx 59,000 words)

Other Swedish Sign Language datasets

Stockholm University

- Dataset of tactile sign language
  - approx. 4 hours of video data (55 media files)
  - 8 deaf-blind signers (age 38-77)
- Dataset of L2 signers
  - 9:30 hours of video data (164 media files)
  - 18 hearing students at SU who are learning SSL as a second language
- Dataset of translations of L1/L2 interpreters from Swedish spoken language to SSL
- Diachronic dataset of language acquisition of deaf children
- Dataset of old L1 signers,
  - 200 signers, older age from 65

Stockholm University

Work resources

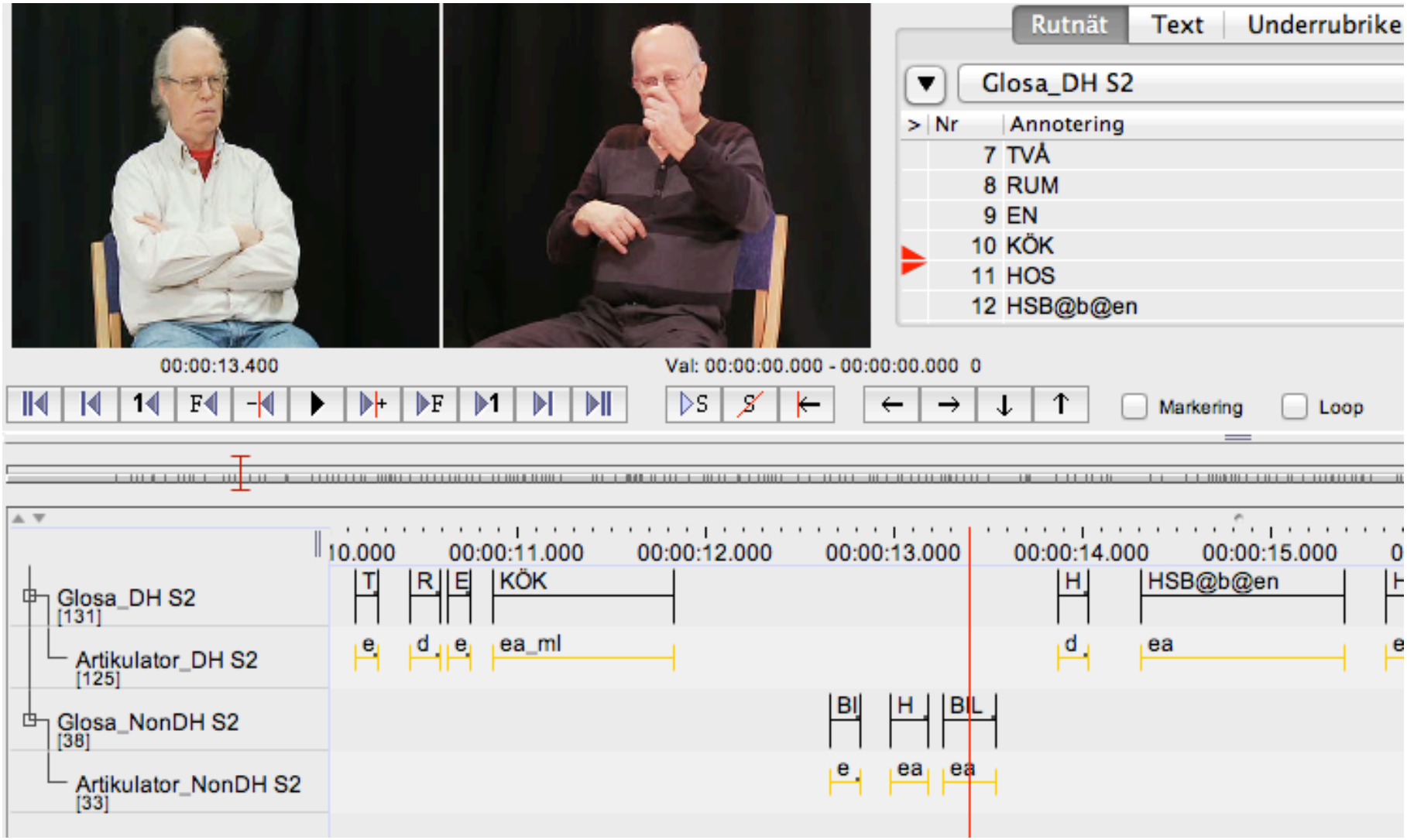
- Limited work resources (only 0.2 full time employment)
- Work resources for Swedish Sign Language dictionary (1.5 full time employment)
  - Started in 1988 and outside the corpus
  - Today, approaching each other with glosses in the dictionary
  - All dictionary entries do not yet have glosses

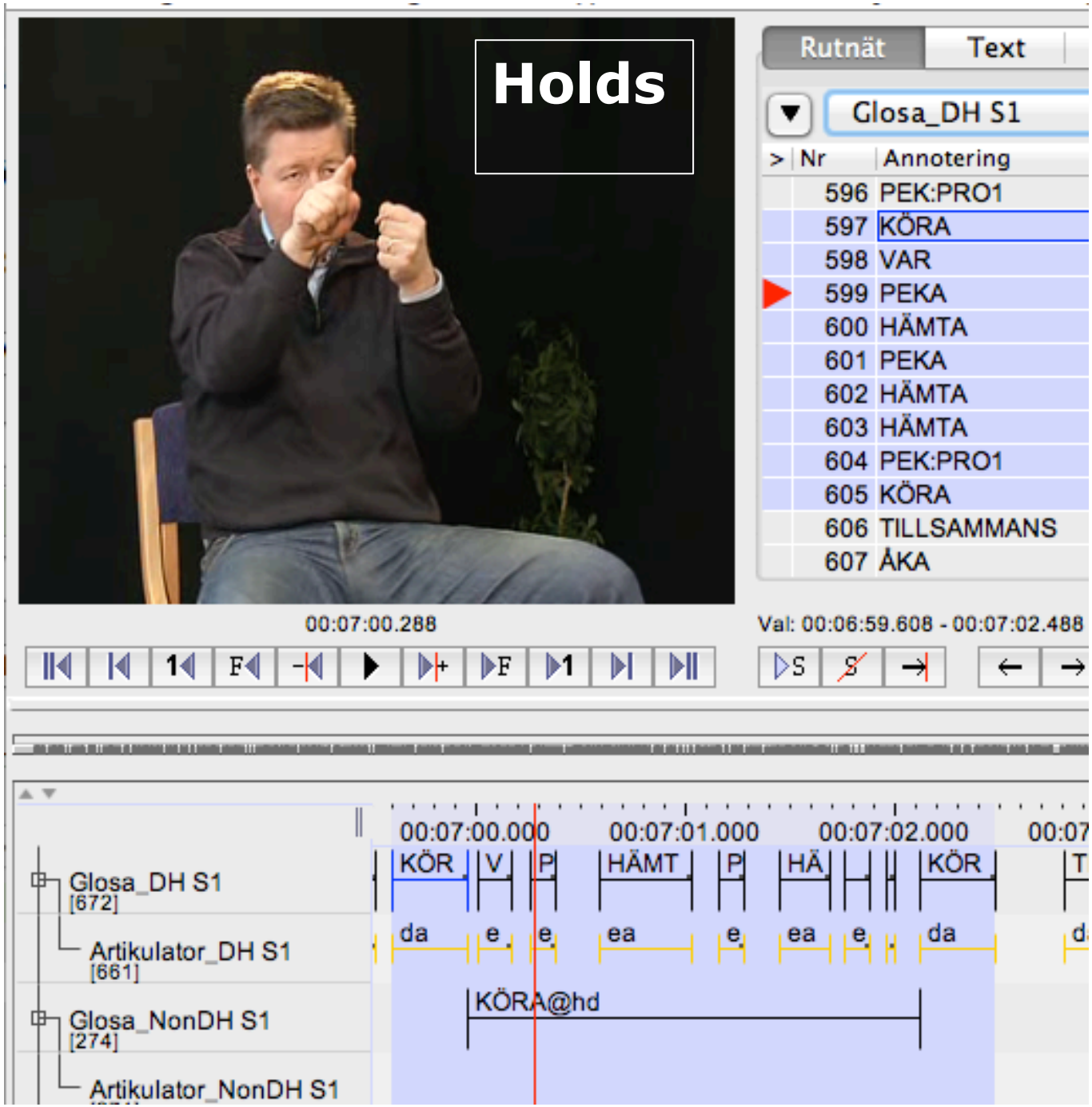
Stockholm University

One sign – one tier
(Basic level)

- Tier Glosa_DH
  - One-handed and two-handed signs
  - Right-handed and left-handed signer
- Tier Glosa_NonDH
  - Buoys
  - Dominance reversals
  - Holds (incl. potential fragment buoys)
  - Overlapping or hold in Glosa_DH

Dominance reversals





Stockholm University

Stockholm University

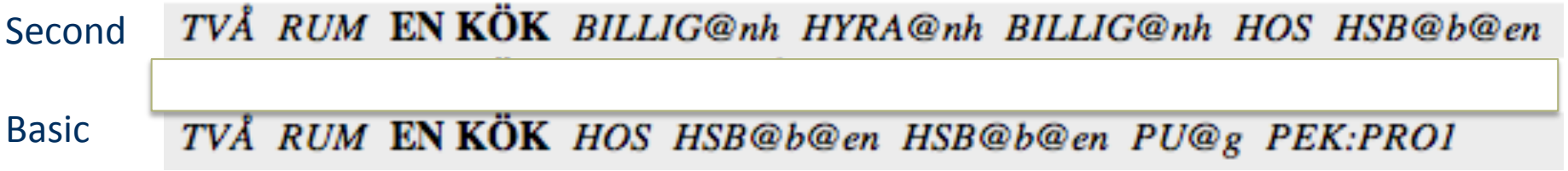
One-handed? Two-handed?

- Two child tiers to Glosa_DH and Glosa_NonDH respectively
  - Articulator_DH
  - Articulator_NonDH

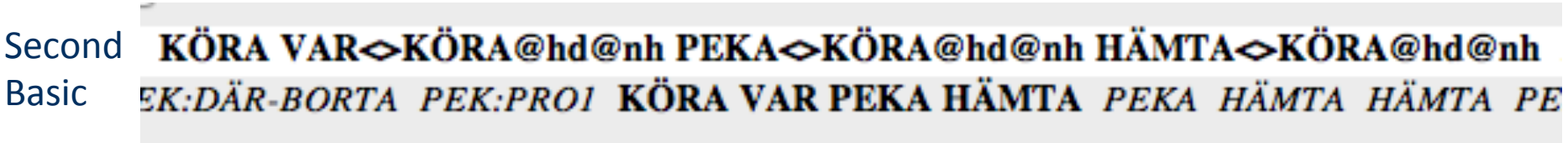
- One-handed
  - ea (single articulator)
- Two-handed
  - ea_ml (single articulator with manual location)
  - da (double articulators)

Dominance reversals and overlaps
(Second level)

- Copy annotations in the tier Glosa_NonDH in the tier Glosa_DH
  - Dominance reversals GLOSS@nh



  - Overlaps GLOSS<>GLOSS@nh



1. General annotation	SSL	BSL	NGT	Discussions/ feedback
1.Basic gloss	SIGN Synonym: (hs) Polyseme: different gloss depending on mouthing No agreement yet wrt lexeme issue in Sweden	SIGN SIGN02	SIGN SIGN-A	Synonym: easier to understand which sign with handshape
2.Two-handed	Parent tier: SIGN Child tier: ea/ea_ml/da	SIGN SIGN	SIGN SIGN	Always did it. Unintuitive with one sign in two tiers
3.Buoys	BUOY:LIST.ONE BUOY:THEME	LBUOY	COUNTING-HAND-1	Prefer buoy
4.Lexical variants	Minimal, but sometimes differentiated with (hs)	SIGN02	SIGN-B	Easier to understand which sign with handshape

1. General annotation	SSL	BSL	NGT	Discussions/ feedback
6.Repetition	If acting like CA: WOLF WOLF WOLF Reduplication or ideophones: WRITE@rd	WOLF WOLF WOLF	WOLF WOLF WOLF	We think that we need to separate two kinds of repetitions
7. Compounds	DEAF^SCHOOL	PARENTS GRAPHIC^ART	PARENTS GRAPHIC-ART	Prefer BSL. Need to distinguish between simple sign and compound sign.
8. Manual negative incorporation	KNOW-NOT UNDERSTAND-NOT	KNOW-NOT	KNOW-NOT	No negative incorporation as number incorporation, but negative component

1. General annotation	SSL	BSL	NGT	Discussions/ feedback
9. Directional verbs	ASK1 Moves towards the signer instead for away in the basic sign RECRUIT1 Moves away from the signer, instead for towards.	ASK TAKE-OVER	ASK:1 1:TAKE-OVER	Numbers only indicate a change in direction compared to the citation form. Direction of backwards verbs is implied.
10. Plurality	SATURDAY+FLERTAL TO-BOOK+FLERTAL '+' indicates sign has two morphemes (except for PEK:PRO.FLERTAL) BOY single movement BOY.PL repetition	CHILD	CHILD.PL	FLERTAL in English would be plural or multiple.

1. General annotation	SSL	BSL	NGT	Discussions/ feedback
11. Numbers	ONE	ONE ONE2	1-A 1-B	Prefer word to digit
12. Number sequences	Same as BSL with carets 2000 in SSL: TWO^ZERO.PLURAL	NINETEEN^EIGHT^NINE	1989	1989 performed with one sign or two?
13. Number incorporation	YEAR+TWO	HOUR-FOUR02	HOUR-4	GLOSS-GLOSS monomorphemic GLOSS+GLOSS dimorphemic, here one for year and one for number
14. Ordinal numbers	ORDER+TWO	FIRST RANKING RANKING02 RANKING02-THREE	1-ORD	

1. General annotation	SSL	BSL	NGT	Discussions/ feedback
15. Sign names	Suffixed LARS-WALLIN@en USA@b@en @en: SN or proper noun @b: FS	SN:FIRST SN:PETER(FS:P-PETER) SN:OSAMA-BINLADEN(BEARD)	*FIRST-LAST	Prefer to read gloss first, then marker
17. Fingerspelling	Suffixed GLOSS@b No info about reduced fingerspelling is. If so, separate tier	FS:WORD FS:WORD(WRD)	#WORD #WRD 'word'	We only gloss target names or concepts, not the phonetic realization of fingerspelling.

1. General annotation	SSL	BSL	NGT	Discussions/ feedback
18. Pointing signs	PT: but Swedish PEK: PEK:PRO1 PEK:PRO for both 2 and 3. PEK>one (index finger at list buoy) PEK>bodyparts	PT:LOC PT:DET PT:PRO1SG Multiple possible functions: PT:LOC/ PT:PRO3SG PT:LBUOY	PT PT:B PT:W PT:1	PT:/PEK: makes it easier to not get caught in pronouns. Proposal to proceed with pronouns. PRO1 is a deictic pointing to oneself, so proposal PT:SELF (cf. Hochgesang 2015). Deictic pointing to the listener or recipient could be PT:REC. Sometimes pointing to an invisible recipient or surrogate could be PT:REC.SURR, etc.

1. General annotation	SSL	BSL	NGT	Discussions/ feedback
19. Classifier/ depicting signs	FIGURE+MOTION (+GROUND)@p Suffixed FIGURE/GROUND: HUMAN-BEING(hs) ENTITY(hs) GRIP(hs) MOTION: MOVE BE-AT PIVOT HANDLE	DSEW(2)-MOVE DSEP(1)-PIVOT DSEW(2)-AT	walks to and fro') PIVOT+1 ('cat's legs move around') AT+2 ('bird is here')	Good idea to minimize the MOTION-component with hyperonymous words such as MOVE, BE-AT. And describe the sign's content in a separate meaning tier.

1. General annotation	SSL	BSL	NGT	Discussions/ feedback
20. Shape constructions	SHAPE(hs) +EXTENSION/FORM	DSS(CYL)	SHAPE +cylinder 'drain pipe'	
21. Type-like classifier/ depicting signs	Annotate as classifier/ depicting sign	DSEW(1-VERT)-MOVE:HUMAN DSEW(FLAT-LATERAL)-AT:VEHICLE	MOVE+1 MOVE+flat	
22. Gestures	GLOSS@g Mostly palm up (PU) Using also CA as alternatives to gesture. THINK@ca (with hand under chin)	G:HOW-STUPID-OF-ME	% HEY	Prefer G to % Some gestures are perceived more as CA when whole body is involved

1. General annotation	SSL	BSL	NGT	Discussions/ feedback
23. Palm up	PU@g We have not yet annotated the exact meaning of gestures. Hope we can change that soon.	G:WELL	PO	Prefer G to without Prefer that we try to annotate what kind of 'gesture' wrt meaning not only form
24. Manual ca	Suffixed @ca POSE@ca	G:CA:HOLD-HANDS-UPIN-FRIGHT	%	Not only manual, but together with non-manual Prefer that we try annotate what kind of 'manual ca'

2. Uncertainties	SSL	BSL	NGT	Discussions
Doubt as to whether the movement is a sign or not	ZZZ@z described in comments	INDECIPHERABLE	±	
Doubt about whether this gloss is chosen correctly	GLOSS@z described the doubt in comments	?GLOSS , or GLOSS1/GLOSS2	?GLOSS	
First annotator doesn't know this sign: it needs to be double-checked by colleagues	GLOSS:(?)@z described in comments	ADD-TO-SIGNBANK(UNKNOWN)	??	

2. Uncertainties	SSL	BSL	NGT	Discussions
Non of us knows this sign	GLOSS:(?)@z	ADD-TO-SIGNBANK(UNKNOWN)	???	
Proposal for a new gloss, to be discussed at the weekly annotation meeting and then added to the SignBank lexikon	GLOSS.NEW:GLOSS@z described new sign in comments	ADD-TO-SIGNBANK(GLOSS)		
New gloss needed, no proposal yet	GLOSS.NEW:(?)@z described new sign in comments	-	$	

2. Uncertainties	SSL	BSL	NGT	Discussions
Invisible, unclear or doubtful which sign it is	ZZZ@z	Options given separated by / ,or INDECIPHERABLE		
False start, but the is recognised as GLOSS	gloss@&	GLOSS(FALSE-START)	˜GLOSS	
False start, not clear what the sign was going to be	zzz@&	INDECIPHERABLE(FALSE-START)	˜	

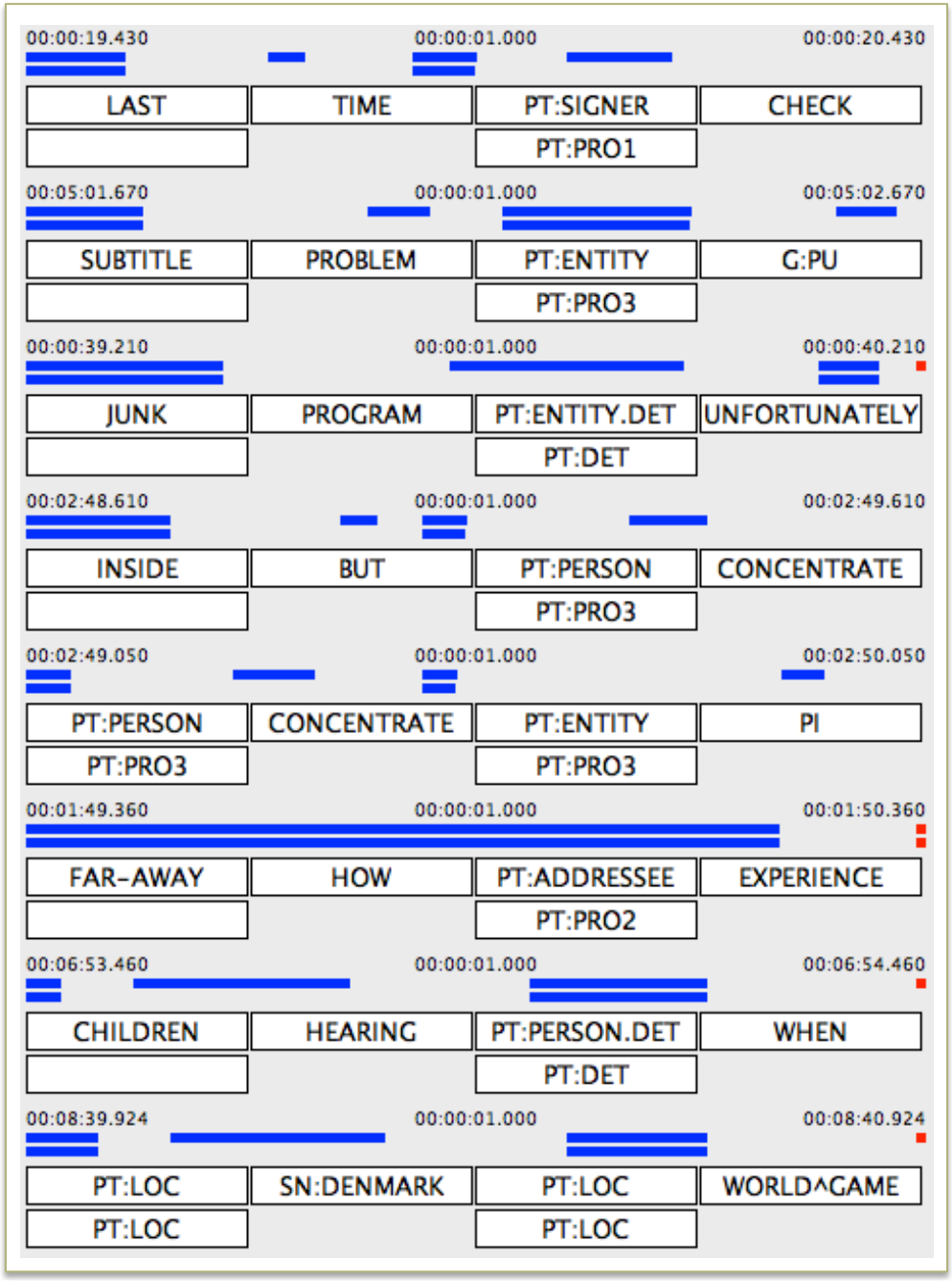
SSL Further annotation	SSL	BSL	NGT	Discussions
Noun classifier sign	@kl TEACHER PERSON@kl			
Fused signs	Asterisk KNOW*NOT DEPEND*ON			
Hesitant, break	tp@& (Thought Pause)			
Hold (incl. potential fragment buoy)	@hd In the tier Glosa_NonDH			
When only non-manual expression or gesture	nms			

Structural differences SSLC

Stockholm University

Annotation file	SSLC01_001	Corpus+Type+Number
Video file	SSLC01_001_S001_b (b=body; f=face; p=pair; tu=top)	Corpus+Type+Number+ Participant_number+View
One annotation file per signer pair	SSLC01_001 is linked to SSLC01_001_S001_b SSLC01_001_S002_b	
Parsing	Small gaps (2 frames) between annotations	

A starting point for discussion on the annotation of pointing signs



PT: or IX:

PT:PERSON
PT:ENTITY
or
PT:ENTITY
for both person and entity

Plurality
PT:SIGNER.PL

Present referents

PT:ADDRESSEE.present
PT:ADDRESSEE>present

PT:LOC.knee
PT:LOC>knee
PT>knee

PT:ADDRESSEE.PL>present