

Introduction

- Lemmatised corpora consist of tokens as instantiations of signs (types).
- Tokens usually count as evidence of the signs' use.
- Corpus may contain instances of sign use that do not reflect the sign use of the person uttering them:
 - particularly citing of signs in metalinguistic discussions
 - malformed signing and slips of the hand
 - copying/repeating signs of the interlocutors or from stimulus.
- Such “non-tokens” should be labelled accordingly to be able to search for them specifically or to exclude them from analysis as suitable.

Examples from DGS Corpus

- Multi-purpose language resource for DGS
 - 330 informants from 13 regions
 - Balanced for age, gender, region
 - 20 elicitation tasks (among others: free conversation, experiences as being deaf, metalinguistic discussion of young vs. old signs)
 - Over 1,150 hrs. of footage
 - About 560 hrs. of filmed signing with estimated 3.5 million tokens
- Lemmatisation
 - Running text (without gaps)
 - Still in progress; nearly 50 hrs. lemmatised (May 2016).

Examples from DGS-Feedback

- Answers and comments
 - Collected via online questionnaire on the use of DGS signs
 - Recorded via webcam (or mobile devices)
- Directly referring to the stimuli (sign form w/wo mouthing, meanings)
- Highly metalinguistic content
- Contain many instances of various kinds of “non-tokens”.

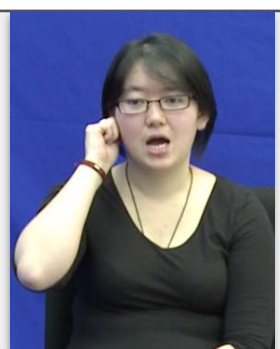
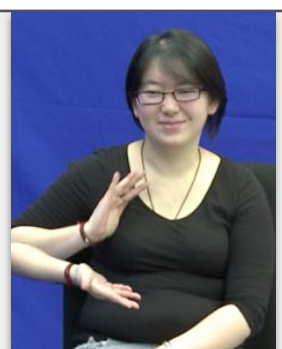
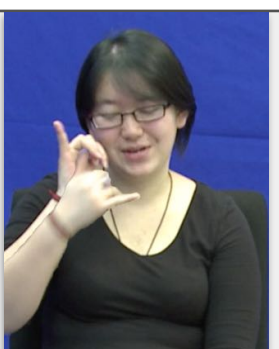
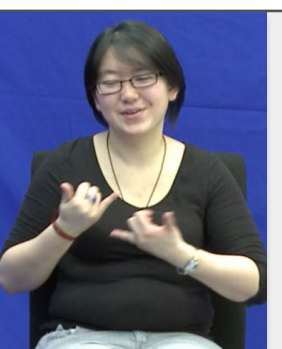
Consequences for Annotation


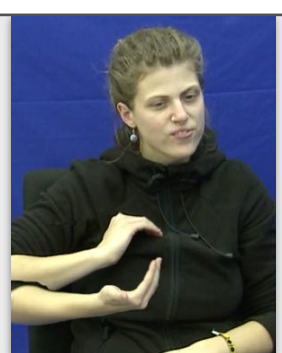
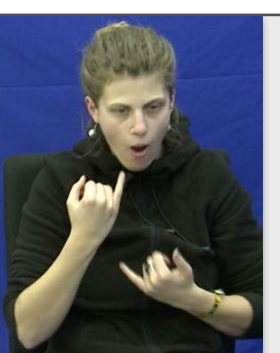
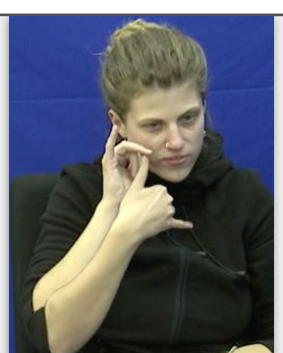
- The kind of “non-token” is a valuable piece of information that needs to be made accessible by some kind of annotation.
- In order to make annotations consistent, these annotations need to be reviewed and groups of “non-tokens” may be refined.

Consequences for Analysis

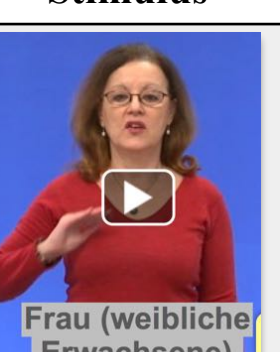
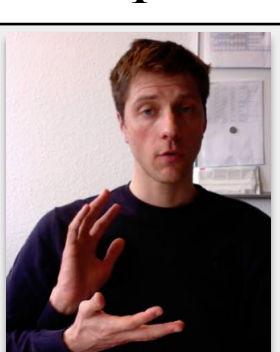



- Making decisions/interpretations of data explicit: Which tokens count as evidence and which do not (“non-tokens”)?
- Retrieving (excluding/including) these tokens for analyses like:
 - Frequency count (count as proof)
 - Count of different signers using a certain sign
 - Research that takes metadata of individual signers such as age, sex, and region into account.

This publication has been produced in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities.



Informant A		A1	A2				A3	A4					A5	A6	A7
															
Gloss		II	WOMAN3B	...			YES1A	WOMAN1	...				WOMAN2	WOMAN4	YES1A
Mouthing			frau					frau					frau		frau
Non-manual								head nod						head nod	
Translation			I use WOMAN3B.					Yes, the sign WOMAN1.					Yes, WOMAN2 and WOMAN4.		
“Non-token”			1a					1b					1b/1d	1b/1d	

Informant B			B1	B2		B3		B4	B5	B6	B7				
															
Gloss		...	WOMAN3B	INDEX1	...	WOMAN1	...	WOMAN4	OR5	WOMAN2	\$GESTURE				
Mouthing		...	frau					frau							
Non-manual			head tilt			head nod			oder	frau					
Translation	How do you sign 'woman'?		You use WOMAN3B.			It is interesting, there is an ongoing discussion. The sign traditionally used everywhere in Germany is WOMAN1,		then new signs evolved here and there: WOMAN4 or WOMAN2.							
“Non-token”			1d			1c		1c		1c					


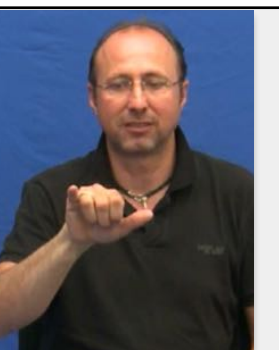
Example 1: DGS Corpus, task *Free Conversation*. A discussion of the different signs for 'woman' and which ones are considered appropriate and should be used and which are not.

	Stimulus	1	2	3	4	5	6	7	8	9	10	11	12	13
														
Gloss		WOMAN1	SIGN1A	II	USE1	NOT1	WOMAN1	YES1A	\$GESTURE	II	USE1	WOMAN2	OR3	WOMAN4
Mouthing		frau	frau	ich	benutze	nicht	busen	ja		ich	benutze	frau	oder	frau
Non-manual		forward lean			headshake									
Translation			I don't use the sign WOMAN1 for 'woman',				but for 'breast'.				I use WOMAN2 or WOMAN4.			
“Non-token”		3/2c					1a					1a		1a

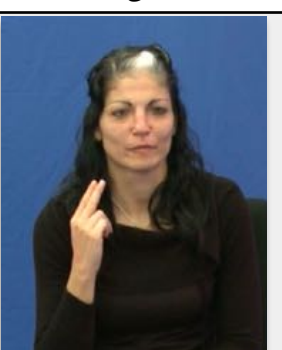
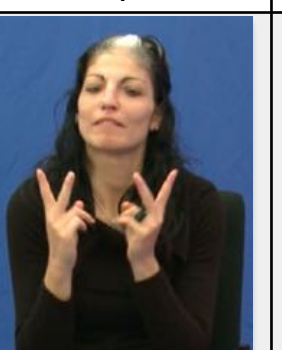
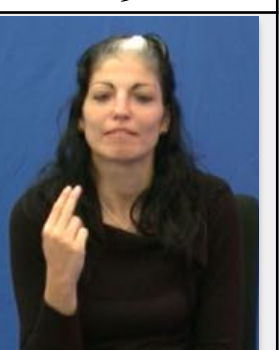
Example 2: Feedback answer to stimulus WOMAN1/'woman' (female adult); re-recorded for anonymisation

	1	2	3	4
				
Gloss	INDEX1	BANANA1A	II	BANANA2
Mouthing		banane		banane
Non-manual		raised eyebrows, head and shoulders forward		
Translation		She signs BANANA1A, I sign BANANA2.		
“Non-token”		1b		1a

Example 3: DGS Corpus, task *Young vs. Old Signs*

	1	2	3	4	5
					
Gloss	ENGLAND3	INDEX1	YES1B		YES-CORRECT
Mouthing	englisch				yes
Non-manual		raised eyebrows		headshake	
Translation		In England they don't sign YES1B, but YES-CORRECT.			
“Non-token”			1a		1e



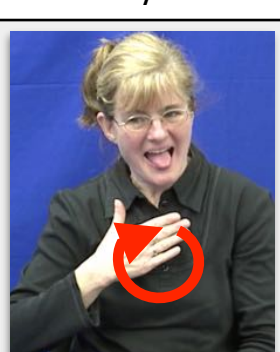
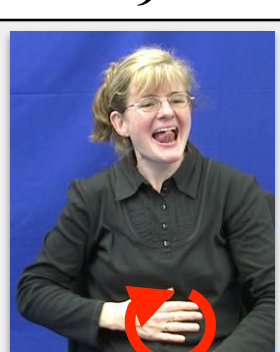

Example 4: DGS Corpus, task *Discussion*

	1	2	3	...	4	5	6	7	8	9
										
Gloss	MOTHER2	SIGN1	IMPORTANT1*	...	SIGN1	NORMAL1	SIGN1	IMPORTANT1	INDEX1	IMPORTANT1*
Mouthing	mutter		wichtig			normal		wichtig		wichtig
Non-manual			head and right shoulder forward							head and shoulders forward
Translation		My mother signs IMPORTANT1*.		...		Normally one signs IMPORTANT1,		but she signs IMPORTANT1*.		
“Non-token”			1g					1a		1g

Example 5: DGS Corpus, task *Experiences as being Deaf*

	Stimulus	1	2	3	4	5	6		7
									
Gloss	CHEESE2A	CHEESE2B	II	CHEESE2A	NOT1	II	CHEESE2B	[hands down]	CHEESE2B-adapted
Mouthing	käse	käse		käse					
Non-manual		raised eyebrows		headshake					
Translation		Concerning the item 'Cheese':		I do not use CHEESE2A.			My sign for 'cheese' is CHEESE2B.		Like this.
“Non-token”				3/2b			1a		5d

Example 6: Feedback answer to stimulus CHEESE2A/'cheese' (food)

	1	2	3	4	5	...	6	7	8	9	10
											
Gloss	TASTY3	ITALY1	BELONG-TO1	SAY1	\$GEST-TASTY	...	INDEX1	TASTY1	II	TASTY*	TASTY3
Mouthing	lecker	italien			lecker			lecker		lecker	lecker
Non-manual								head tilt [smile]	[smile]	[smile]	[smile]
Translation		For 'tasty' in Italy they use this gesture.					You sign TASTY1,		I sign TASTY*, TASTY3.		
“Non-token”					4		1d		5b		1a

Example 7: DGS Corpus, task *Free Conversation*

	Stimulus		1	2	3	4	5	
								
Gloss	HUNGER1	...	12	SIGN	HUNGER5	HUNGER5*	HUNGER5	...
Mouthing	hunger				hunger		hunger	
Non-manual					erect stance			
Translation		And 'hunger' [fingerspelled],			I sign HUNGER5			
“Non-token”					1a	5d	1a	

Example 8: Feedback answer to stimulus HUNGER1/'hunger' (bodily need for food)

Kinds of “Non-tokens”

- 1 Metalinguistic reference to existing signs
 - a DGS sign also used by the signer
 - b Existing DGS sign that is normally not used by the signer
 - c Unclear cases (not clear whether a or b)
 - d Reference to a sign from his/her interlocutor
 - e Reference to a sign of a foreign sign language
 - f Reference to a sign that the signer has been using in the past, but does not use anymore
 - g Reference to an idiosyncratic sign use
- 2 “Un-tokens” (often used contrastively): Sign shown and metalinguistically explicitly negated (by headshake, facial expression and/or negation sign) to indicate that the sign, form or use shown is not used by the signer
 - a “Non-sign”: negation of whole sign
 - b “Non-form”: negation of this particular execution of form
 - c “Non-use”: negation of use with regard to a particular meaning
- 3 Copy of stimulus: Reference to stimulus by copy of form
- 4 “Non-signs” (of SL in concern): Reference to gestures of hearing persons
- 5 Malformed or deviated sign forms
 - a False start and aborted articulation
 - b Slip of the hand
 - c Trying out or searching for signs
 - d Adapted articulation (due to limited view of webcam/recording device)

Poster presented at the
7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining, at the 10th Language Resources and Evaluation Conference – LREC 2016, in Portorož, Slovenia. May 28, 2016.