



section
for sign
linguistics

London, 30 March 2015

PJM CORPUS ANNOTATION GUIDELINES

Paweł Rutkowski, Joanna Filipczak, Anna Kuder
Section for Sign Linguistics
University of Warsaw

PJM Corpus Project

- Developed since 2011.
- 2010: a grant awarded by the Foundation for Polish Science within the FOCUS program (equipment, corpus architecture, beginning of data collection, preliminary annotations)



- 2011: a grant awarded by the National Science Centre within the HARMONIA program (dictionary, iconicity)



PJM Corpus Project

Currently:

- 2014-2019: a grant awarded within the National Programme for the Development of Humanities of the Polish Ministry of Science and Higher Education (multi-layer annotation)



NARODOWY PROGRAM
ROZWOJU HUMANISTYKI

- Collaboration with prof. Trevor Johnston (as an external advisor). THANKS, TREVOR! 😊

Corpus statistics (March 2015)

173 652
revised
tags

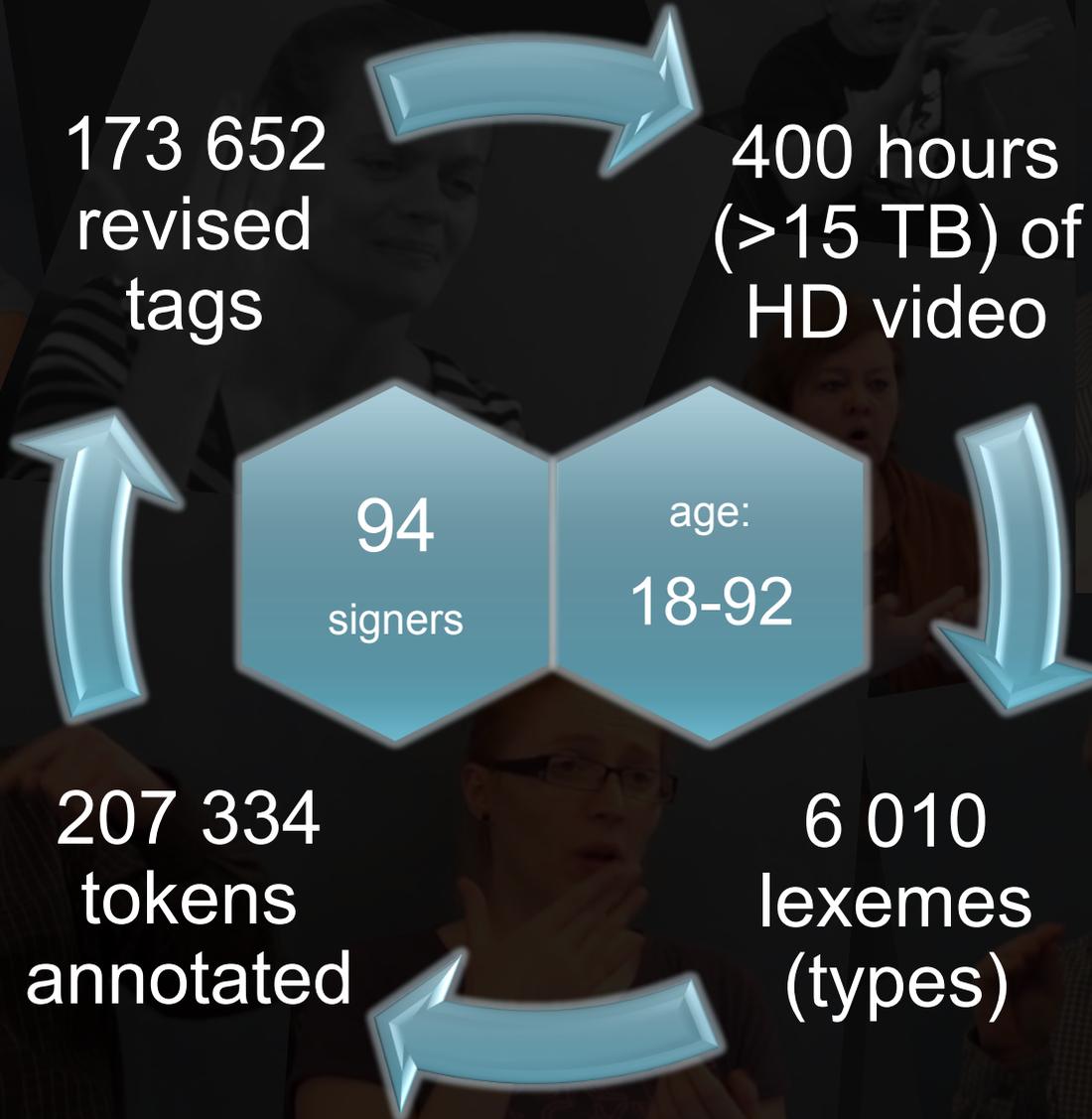
400 hours
(>15 TB) of
HD video

94
signers

age:
18-92

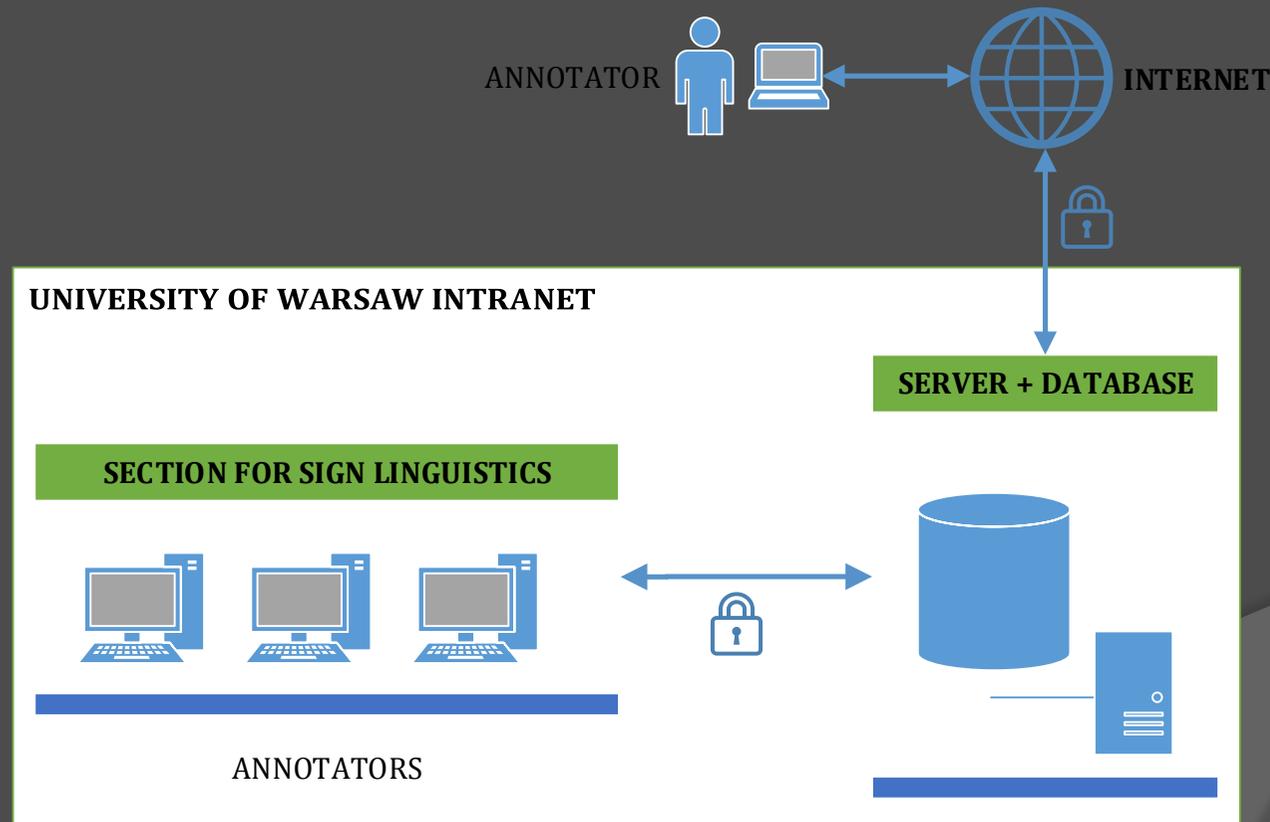
207 334
tokens
annotated

6 010
lexemes
(types)

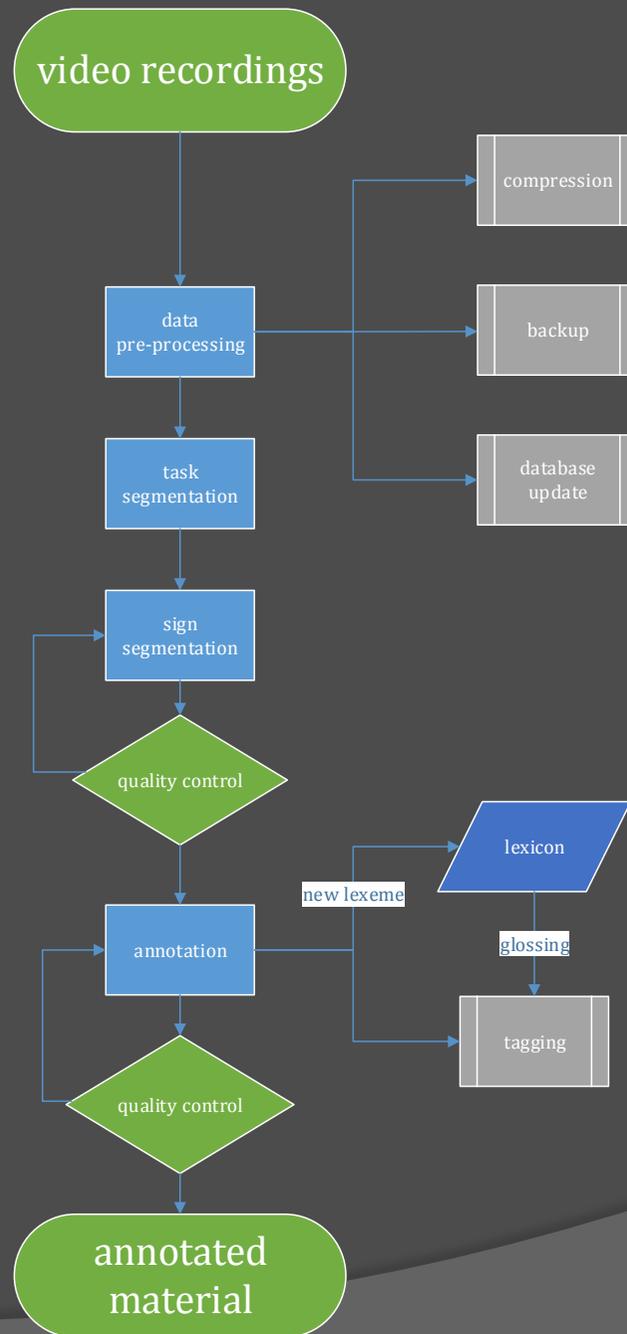


Centralized multi-person annotation system

ACCESS TO THE SERVER AND DATABASE



Workflow



18 from K03BF01-21



02:48:54:20 M! →M >|<

Segment Name:
18

Theme:

Transcript:
K03BF01-21

Timecodes^	nr_zad...	elicytacja	glosa	glosa_...
02:48:51:03 02:48:51:19			RÓZNY/RÓZNIE/RÓZNICA/INNY	
02:48:51:19 02:48:52:00			WSKAZ: Z (WSZYSTKIE KIERUNKI/STRZELAĆ/JEŻ	
02:48:52:00 02:48:53:05				
02:48:53:05 02:48:54:09			DENTYSTA 1.2 P:AX;L:Ø	
02:48:54:09 02:48:54:21			SPRAWDZAĆ/SPRAWDZIĆ/KONTROLA 1.1 P:AX	
02:48:54:21 02:48:55:01			LEPIEJ 1.3 P:E;L:Ø	
02:48:55:01 02:48:55:11			WAŻNY P:W;L:W	
02:48:55:11 02:48:55:18			ZĄB P:Z;L:Ø (ZĘBY)	
02:48:55:18 02:48:55:22			ŁADNY 1.1 P:5A;L:Ø	
02:48:55:22 02:48:56:03			BIAŁY/BIAŁYSTOK/BIAŁORUŚ/SOBOTA P:B;L:Ø	
02:48:56:03 02:48:56:09			ZĄB P:Z;L:Ø (ZĘBY)	
02:48:56:09 02:48:56:16			MÓWIĆ P:Z;L:Ø (POWIEDZIEĆ/KAZAĆ/ZLECIEĆ/Z	
02:48:56:16 02:48:57:05			SPRAWDZAĆ/SPRAWDZIĆ/KONTROLA 1.1 P:AX	
02:48:57:05 02:48:57:21			ZĄB P:Z;L:Ø (ZĘBY)	
02:48:57:21 02:48:58:17			REGULARNY 1.1 P:O;L:B	
02:48:58:17 02:48:59:03			ROK P:1;L:Ø	
02:48:59:03 02:48:59:12			PÓŁ P:Z;L:Z > DO POPRAWY (ROZDZIELAĆ)	
02:48:59:12 02:48:59:24			ROK/LATA/ROZCZNIKA/IDENTYF: (ROZCAN) P:1	
02:48:59:24 02:49:00:05			WSKAZ: Z (WSZYSTKIE KIERUNKI/STRZELAĆ/JEŻ	
02:49:00:05 02:49:00:19			ZOSTAWIĆ P:5;L:5	
02:49:00:19 02:49:03:11				
02:49:03:11 02:49:04:07			LEKARZ/DOKTOR/AKURAT P:O;L:B	

Tiers

TIMECODES GLOSSES HAMNOSYS TRANSLATION TAGS



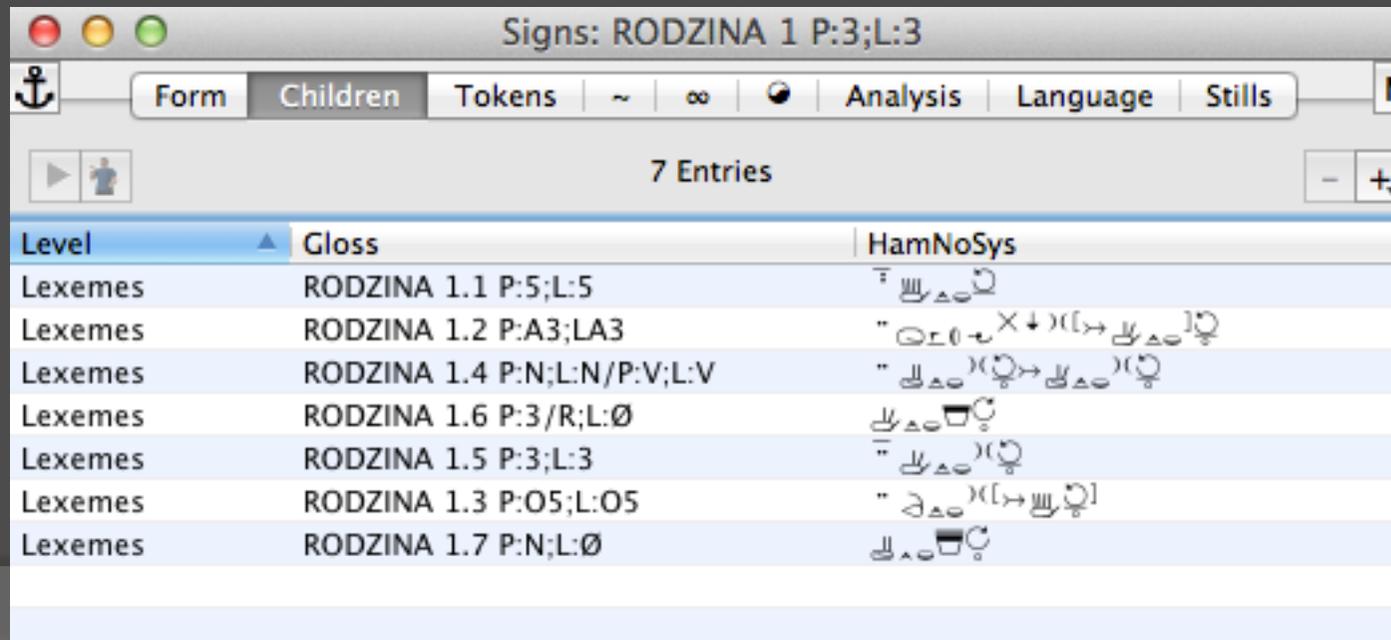
Timecodes▲	glosa	glosa_...	HamNoSys	HamNoSys_dew	tlumaczenie	PoS	NMNS-INTON	NMNS-BODY	MOUTHING	PoS_glosa_druga...	REPETY
00:28:43:05 00:28:44:10	WSKAZ: Z (JA)		↓ x 0 ̄ X		tu(tutaj), coś	IX					
00:28:44:10 00:28:45:07	###										
00:28:45:07 00:28:45:20	CZUĆ/FORMA/OBAWIA		̄ ̄ < ̄ ̄ ± X +			V					
00:28:45:20 00:28:47:11	@				hej(woła kog	PHAT					
00:28:47:11 00:28:48:05	CZUĆ/FORMA/OBAWIA		̄ ̄ < ̄ ̄ ± X +			V		H_nod	czułam		
00:28:48:05 00:28:49:16	###										
00:28:49:16 00:28:50:13	TROCHEŃ/ZARAZ/CHWI		̄ ̄ a ̄ ̄ ̄ → ̄ ̄		Za chwilę. Ta	ADV		H_nod	chwilę		
00:28:50:13 00:28:54:00	###							H_nod			
00:28:54:00 00:28:54:15	CZUĆ/FORMA/OBAWIA		̄ ̄ < ̄ ̄ ± X +			V			czuje		
00:28:54:15 00:28:59:17	###										
00:28:59:17 00:29:00:11	MINUS/KRESKA P:Z:L:Ø				Minus(suger	N			minus		
00:29:00:11 00:29:01:22	###										
00:29:01:22 00:29:02:14	MUSIEĆ/POWINNO/TR				Muszę, Tak.	NorV			mu-		
00:29:02:14 00:29:12:20	###										
00:29:12:20 00:29:13:21	TAK P:3:L:Ø >TAK /OK		̄ ̄ 2 3 a ̄ ̄ ̄ → ̄ ̄					H_nod			
00:29:13:21	WSKA7- 7 (WSZYSTKIF		̄ ̄ a ̄ ̄ ±		To(światło) n	IX					

Basic gloss

- ⦿ Annotation glosses correspond to lemmas.
- ⦿ Two-level annotation for types (citation forms) and variants.
- ⦿ Distinction into types is always based on the manual layer.
- ⦿ Variants (=subtypes) are sign forms that differ in no more than one of the following features: handshape, location, movement, orientation.

Subtypes

- Related two-handed and one-handed variants are treated as subtypes of the same type.
- Change in one parameter gives a new variant within a given type.



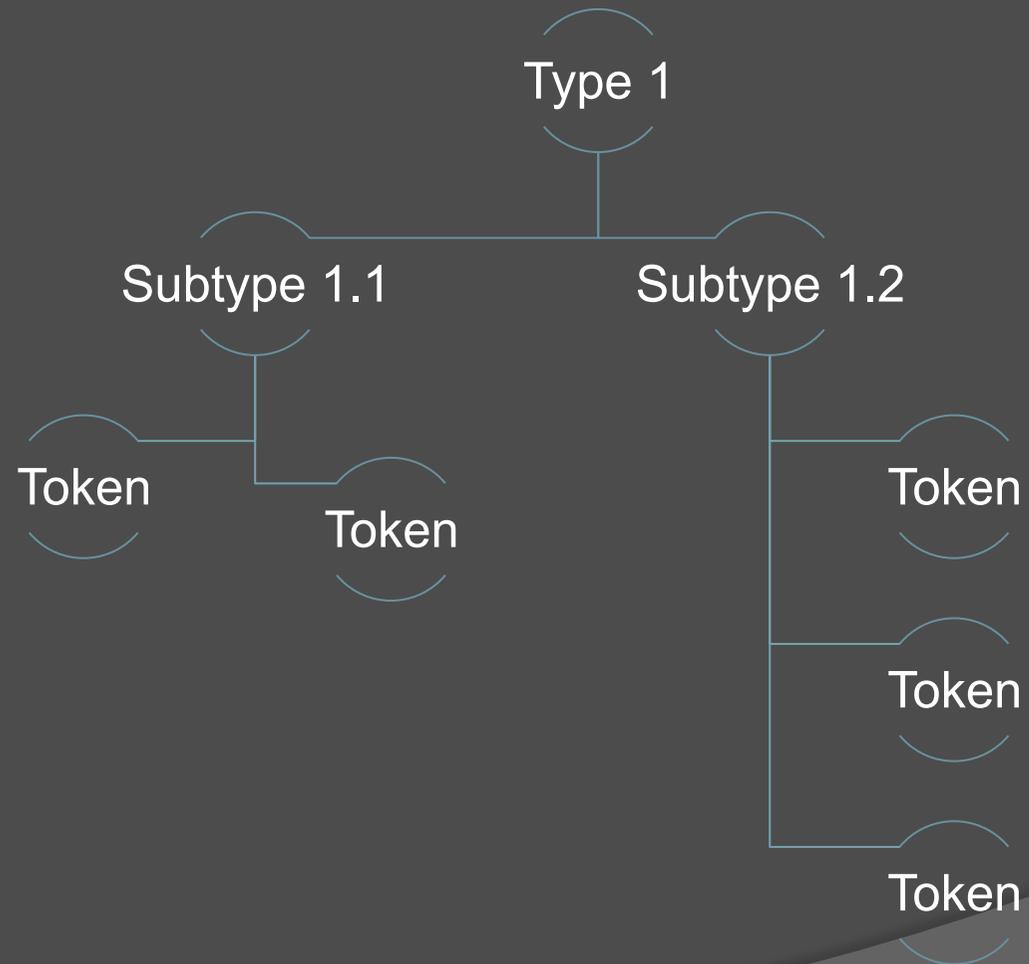
Signs: RODZINA 1 P:3;L:3

Form Children Tokens ~ ∞ Analysis Language Stills

7 Entries

Level	Gloss	HamNoSys
Lexemes	RODZINA 1.1 P:5;L:5	⌘ ⌘ ⌘ ⌘ ⌘
Lexemes	RODZINA 1.2 P:A3;LA3	" ⌘ ⌘ ⌘ X +) ([⌘ ⌘ ⌘ ⌘ ⌘
Lexemes	RODZINA 1.4 P:N;L:N/P:V;L:V	" ⌘ ⌘ ⌘ ⌘ ⌘ ⌘ ⌘ ⌘ ⌘ ⌘
Lexemes	RODZINA 1.6 P:3/R;L:∅	⌘ ⌘ ⌘ ⌘ ⌘
Lexemes	RODZINA 1.5 P:3;L:3	" ⌘ ⌘ ⌘ ⌘ ⌘
Lexemes	RODZINA 1.3 P:O5;L:O5	" ⌘ ⌘ ⌘ ⌘ ⌘ ⌘ ⌘ ⌘ ⌘ ⌘
Lexemes	RODZINA 1.7 P:N;L:∅	⌘ ⌘ ⌘ ⌘ ⌘

Two-level annotation



Conventions for gloss naming

- Each gloss is followed by a number and a handshape marker for the right and left hand (based on PJM fingerspelling). Additional info in brackets if more than one sign employ the same handshape.
- Variants have the same naming conventions.
- TYPE:

MAMA 1 P:N;L:Ø

SUBTYPE (variants):

MAMA 1.1 P:N;L:Ø (CHEEK)

MAMA 1.2 P:N;L:Ø (CHEEK-CHIN)

MAMA 1.3 P:Z;L:Ø

MAMA 1.4 P:5;L:Ø

No variants?

- A basic gloss has a subtype even when it has only one articulatory variant (in such cases those forms are identical).

The screenshot shows two overlapping windows from a software application. The top window is titled "Signs: ADWOKAT 1 P:A5;L:A5" and has tabs for "Form", "Children", "Tokens", "Analysis", "Language", and "Stills". The "Form" tab is active, showing a "Gloss:" field with the text "ADWOKAT 1 P:A5;L:A5". Below this is a "HamNoSys:" field containing the representation "ʔ Δ ʌ → ʔ +". There is also a "Mouth:" field and a "Composed of:" section with empty input boxes.

The bottom window is also titled "Signs: ADWOKAT 1 P:A5;L:A5" and has the same tabs. The "Children" tab is active, showing a table with "1 Entries". The table has columns for "Level", "Gloss", and "HamNoSys".

Level	Gloss	HamNoSys
Lexemes	ADWOKAT 1.1 P:A5;L:A5	ʔ Δ ʌ → ʔ +

Two-handed signs

- Two handed signs are glossed on the dominant-hand tier only.
- The dominant-hand tier is the tier for the preference hand of the signer (but for the time being we only have right-handed signers in the corpus).

18 from K05AF13-26

Timecodes^	nr_zad...	elicytacja	glosa	glosa_druga_reka	tumac...	komen...	segme...	HamN...
00:31:20:19 00:31:52:20								
00:31:52:20 00:31:53:12			WSKAZ: Z (RUCH PO	WSKAZ: Z (WSZYSTKIE KIERL				
00:31:53:12 00:31:53:22			WYBIERAĆ/WYBÓR/W					
00:31:53:22 00:31:54:15			OBOJĘTNIE P:O:L:Ø					
00:31:54:15 00:31:54:23								
00:31:54:23 00:31:55:16			NUM: DRUGI 1.1 P:V					
00:31:55:16 00:31:57:17								
00:31:57:17 00:31:58:15			LEKARZ/DOKTOR/A					
00:31:58:15 00:32:00:10			\$:KL: A+A (TRZYMA					
00:32:00:10 00:32:01:00			BADAĆ/BADANIE 2.	\$:KL: A+A (TRZYMAĆ+(ROW				
00:32:01:00 00:32:01:24			\$:KL: E+B (BADAĆ D					
00:32:01:24 00:32:03:15			WIDZIEĆ/PATRZEĆ 1					
00:32:03:15 00:32:05:08			\$:KL: E+B (BADAĆ D	\$:KL: A+A (TRZYMAĆ+(ROW				
00:32:05:08 00:32:10:02								
00:32:10:02 00:32:11:00			APTEKA 1.1 P:A:L:B					
00:32:11:00 00:32:11:14								
00:32:11:14 00:32:11:20			DZIEWCZYNA/KOBIE					
00:32:11:20 00:32:13:22			#???					
00:32:13:22 00:32:15:00			PYTAĆ P:O:L:O					
00:32:15:00 00:32:15:21								
00:32:15:21 00:32:16:16			PUDEŁKO/PACZKA/K					
00:32:16:16 00:32:17:02			TABLETKA/LEKARSTV					
00:32:17:02 00:32:17:21			PYTAĆ P:O:L:O					

Tiers

GLOSS FOR DOMINANT AND NONDOMINANT HAND



Timecodes^	glosa	glosa_...	HamNoSys	HamNoSys_dew	tlumaczenie	PoS	NMNS-INTON	NMNS-BODY	MOUTHING	PoS_glosa_druga...	REPETY
00:28:43:05 00:28:44:10	WSKAZ: Z (JA)		↓ x 0 ̄ X		tu(tutaj), coś	IX					
00:28:44:10 00:28:45:07	###										
00:28:45:07 00:28:45:20	CZUĆ/FORMA/OBAWIA		̄ ̄ < ̄ ̄ ± X +			V					
00:28:45:20 00:28:47:11	@				hej(woła kog	PHAT					
00:28:47:11 00:28:48:05	CZUĆ/FORMA/OBAWIA		̄ ̄ < ̄ ̄ ± X +			V		H_nod	czułam		
00:28:48:05 00:28:49:16	###										
00:28:49:16 00:28:50:13	TROCHEŃ/ZARAZ/CHWI		̄ ̄ a ̄ ̄ ̄ → ̄ ̄		Za chwilę. Ta	ADV		H_nod	chwilę		
00:28:50:13 00:28:54:00	###							H_nod			
00:28:54:00 00:28:54:15	CZUĆ/FORMA/OBAWIA		̄ ̄ < ̄ ̄ ± X +			V			czuje		
00:28:54:15 00:28:59:17	###										
00:28:59:17 00:29:00:11	MINUS/KRESKA P:Z;L:Ø				Minus(suger	N			minus		
00:29:00:11 00:29:01:22	###										
00:29:01:22 00:29:02:14	MUSIEĆ/POWINNO/TR				Muszę, Tak.	NorV			mu-		
00:29:02:14 00:29:12:20	###										
00:29:12:20 00:29:13:21	TAK P:3;L:Ø >TAK /OK		̄ ̄ 2 3 a ̄ ̄ ̄ → ̄ ̄					H_nod			
00:29:13:21	WSKA7- 7 (WSZYSTKIF		̄ ̄ a ̄ ̄ ±		To(światło) n	IX					

Two-handed signs

- Change in hand dominance is not tagged. It is assumed when glosses appear only in the tier for the non-dominant hand and the tier for the dominant hand is empty.

Timecodes	glosa	glosa_druga_reka	gramatyka	tlu
02:18:42:18 02:18:56:01				
02:18:56:01 02:18:56:08	MOŻNA/MOŻE/MOŻ			
02:18:56:08 02:18:57:04	WOJSKO P:I;L:Ø/Q: I			
02:18:57:04 02:18:57:15	NUM: JEDEN P:Z;L:Ø			
02:18:57:15 02:18:58:05	WOJSKO P:I;L:Ø/Q: I			
02:18:58:05 02:18:58:20	POCZEKAĆ P:B;L:Ø /			
02:18:58:20 02:18:59:14				
02:18:59:14 02:18:59:23		POTEM P:B;L:Ø		
02:18:59:23 02:19:00:10		NASTĘPNY 2.5 P:B;L:		
02:19:00:10 02:19:03:08				
02:19:03:08 02:19:03:13	TAK-SAMO/TEŻ/RÓ			
02:19:03:13 02:19:03:19	MIEĆ 2.2 P:B;L:Ø (D	WSKAZ: Z (WSZYSTKI		
02:19:03:19 02:19:03:24	NIE-DAWNO/POPRZ			
02:19:03:24 02:19:04:11	S:KL: Z (IŚĆ/JEŹDZIC	WOLNY/FERIE/ODPO		
02:19:04:11 02:19:04:23	MINAĆ/MIJAĆ_SIE 1.			
02:19:04:23 02:19:05:06	POCZEKAĆ P:B;L:B /	POCZEKAĆ P:B;L:B /S		
02:19:05:06 02:19:05:13	TROCHE/ZARAZ/CH			

Simultaneity

- When the second hand articulates a meaningful sign or a classifier When a signer signs two different signs at the same time (not a single two-handed sign).
- Perseveration is annotated by extending duration of a given gloss.

18 from K05AF13-26

Timecodes^	nr_zad...	elicytacja	glosa	glosa_druga_reka	tlumac...	komen...	segme...	H
00:31:20:19 00:31:52:20								
00:31:52:20 00:31:53:12			WSKAZ: Z (RUCH PO	WSKAZ: Z (WSZYSTKIE KIERU				
00:31:53:12 00:31:53:22			WYBIERAĆ/WYBÓR/W					
00:31:53:22 00:31:54:15			OBOJĘTNIE P:O;L:Ø					
00:31:54:15 00:31:54:23								
00:31:54:23 00:31:55:16			NUM: DRUGI 1.1 P:V					
00:31:55:16 00:31:57:17								
00:31:57:17 00:31:58:15			LEKARZ/DOKTOR/A					
00:31:58:15 00:32:00:10			\$:KL: A+A (TRZYMA					
00:32:00:10 00:32:01:00			BADAĆ/BADANIE 2.	\$:KL: A+A (TRZYMAĆ+(ROW				
00:32:01:00 00:32:01:24			\$:KL: E+B (BADAĆ D					
00:32:01:24 00:32:03:15			WIDZIEĆ/PATRZEĆ 1					
00:32:03:15 00:32:05:08			\$:KL: E+B (BADAĆ D	\$:KL: A+A (TRZYMAĆ+(ROW				
00:32:05:08 00:32:10:02								
00:32:10:02 00:32:11:00			APTEKA 1.1 P:A;L:B					
00:32:11:00 00:32:11:14								
00:32:11:14 00:32:11:20			DZIEWCZYNA/KOBIE					
00:32:11:20 00:32:13:22			#???					
00:32:13:22 00:32:15:00			PYTAĆ P:O;L:O					
00:32:15:00 00:32:15:21								
00:32:15:21 00:32:16:16			PUDEŁKO/PACZKA/K					
00:32:16:16 00:32:17:02			TABLETKA/LEKARSTV					
00:32:17:02 00:32:17:21			PYTAĆ P:O;L:O					

Buoys

- List buoys are annotated with special glosses.
- Other buoys are not glossed in a special way, instead, non-dominant hand glosses are extended in duration.

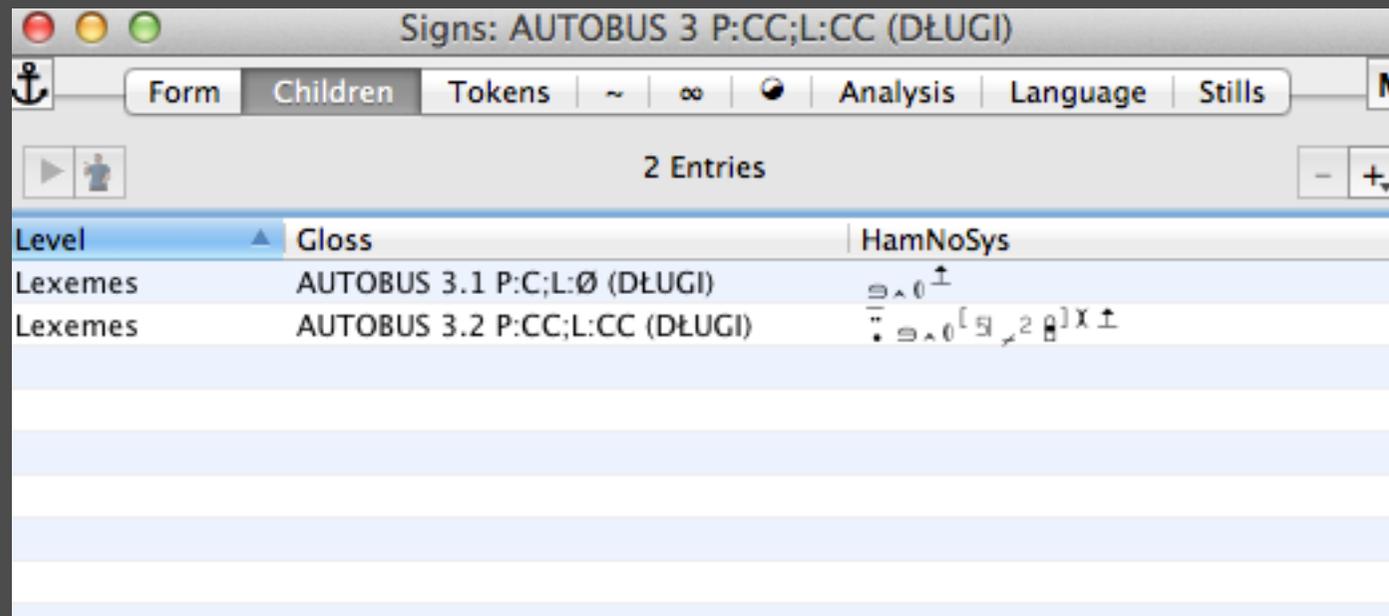
The screenshot shows a software window titled "Signs: PIĄTY-PUNKT 1 P:Z;L:5". The window has a menu bar with "Form", "Children", "Tokens", "Analysis", "Language", and "Stills". Below the menu bar, there are several input fields and sections:

- Gloss:** A text field containing "PIĄTY-PUNKT 1 P:Z;L:5".
- HamNoSys:** A text field containing "[ǂ 1 1 0 4 ǂ 1 1 0] 5 ǂ X".
- Mouth:** An empty text field.
- Composed of:** A section with a checkbox and two empty text fields with double arrows between them.
- Description:** A large empty text area.
- Parent:** A section with a "Gloss:" label and an empty text field.

At the bottom of the window, there is a footer: "2012-11-06 (Michał Grojs), 2014-05-16 (Asia Filipczak)".

Phonological variants

- Phonological variants are distinguished on the subtype level with additional numbers (1.1, 1.2/2.1, 2.2/3.1).



The screenshot shows a software window titled "Signs: AUTOBUS 3 P:CC;L:CC (DŁUGI)". The window has several tabs: "Form", "Children", "Tokens", "Analysis", "Language", and "Stills". Below the tabs, there is a section labeled "2 Entries" with a play button icon and a person icon. The main content is a table with three columns: "Level", "Gloss", and "HamNoSys".

Level	Gloss	HamNoSys
Lexemes	AUTOBUS 3.1 P:C;L:Ø (DŁUGI)	əˈbʊs [±]
Lexemes	AUTOBUS 3.2 P:CC;L:CC (DŁUGI)	ˈɔːəˈbʊs [±] [ɔːˈbʊs [±]]

Repetition

- ⦿ Repetition interpreted by the annotator as a grammatical form (as plural/aspectual modification) is not annotated separately (but is annotated in HamNoSys transcription for the token).
- ⦿ Signs that are repeated by the signer without the intention of grammatical repetition are annotated separately.

Compounds

- ◉ No separate symbol for possible compounds. Information added after the articulation info as SIGN+SIGN.

Level	Gloss	HamNoSys
Lexemes	KOŚCIÓŁ 1.1 P:ZB;L:ZB (KRZYŻ+DOM)	" ɔ̌ r ɔ̌ ɔ̌ x̌ → ɔ̌ r ɔ̌ x̌ ɔ̌ *
Lexemes	KOŚCIÓŁ 1.2 P:BZ;L:BZ (DOM+KRZYŻ)	" ɔ̌ r ɔ̌ x̌ ɔ̌ * → ɔ̌ ɔ̌ ɔ̌ x̌

Directional verbs

- Directionality info: the subtype level.

Gloss:

POWIEDZIEĆ/PYTAĆ 1 P:B;L:Ø / P:M;L:Ø (POKLEPAĆ+RAMIE)

HamNoSys:

Signs: POWIEDZIEĆ/PYTAĆ 1 P:B;L:Ø / P:M;L:Ø (POKLEPAĆ+RAMIE)

Form Children Tokens ~ ∞ Analysis Language Stills M

5 Entries

Level	Gloss
Lexemes	POWIEDZIEĆ/PYTAĆ (MI) 1.1 P:B;L:Ø / P:M;L:Ø (POKLEPAĆ+RAMIE)
Lexemes	POWIEDZIEĆ/PYTAĆ (KOMUŚ/TOBIE) 1.2 P:B;L:Ø / P:M;L:Ø (POKLEPAĆ+RAMIE)
Lexemes	POWIEDZIEĆ/PYTAĆ (MI) 1.3 P:B;L:B / P:M;L:Ø (POKLEPAĆ+RAMIE)
Lexemes	POWIEDZIEĆ/PYTAĆ (MIEJU OSOBOM) 1.4 P:B;L:Ø / P:M;L:Ø (POKLEPAĆ+RAMIE)
Lexemes	POWIEDZIEĆ (KOMUŚ) 1.5 P:B;L:B

Annotation strategies

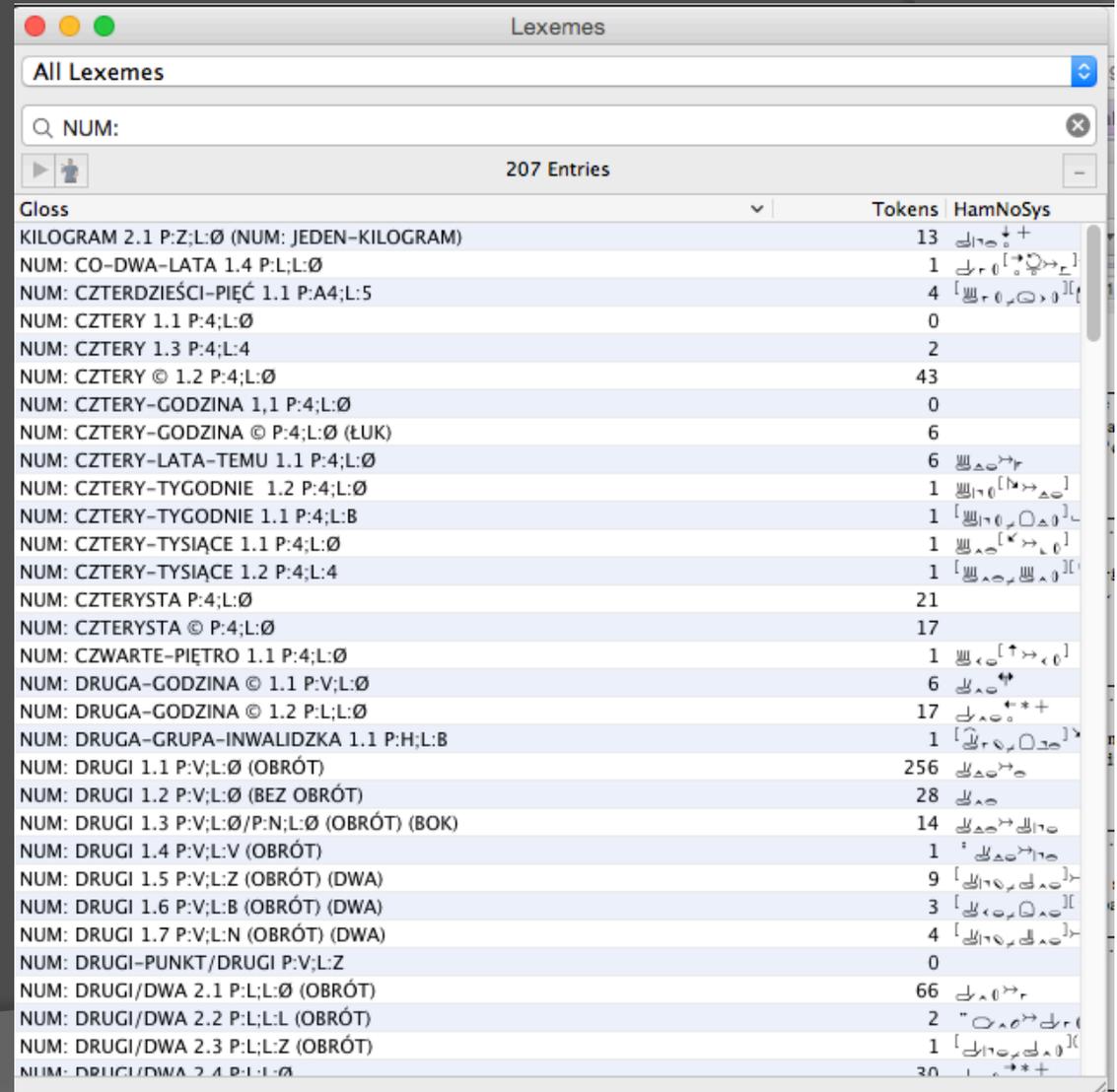
- ① Directionality is encoded in the name on the gloss at the subtype level. This solution was suggested by our Deaf annotators.
- ① Additional tags concerning directionality will be encoded on a separate tier.

Plurality

- ⦿ Grammatical/modification info is annotated on the token level (no additional tier needed).

Numbers

- Glosses for numbers receive the prefix NUM: and are written in words.



Lexemes

All Lexemes

Q NUM:

207 Entries

Gloss	Tokens	HamNoSys
KILOGRAM 2.1 P:Z;L:Ø (NUM: JEDEN-KILOGRAM)	13	ك١٠٠٠
NUM: CO-DWA-LATA 1.4 P:L;L:Ø	1	ك١٠٠٠
NUM: CZTERDZIEŚCI-PIĘĆ 1.1 P:A4;L:5	4	ك١٠٠٠
NUM: CZTERY 1.1 P:4;L:Ø	0	
NUM: CZTERY 1.3 P:4;L:4	2	
NUM: CZTERY © 1.2 P:4;L:Ø	43	
NUM: CZTERY-GODZINA 1,1 P:4;L:Ø	0	
NUM: CZTERY-GODZINA © P:4;L:Ø (ŁUK)	6	
NUM: CZTERY-LATA-TEMU 1.1 P:4;L:Ø	6	ك١٠٠٠
NUM: CZTERY-TYGODNIE 1.2 P:4;L:Ø	1	ك١٠٠٠
NUM: CZTERY-TYGODNIE 1.1 P:4;L:B	1	ك١٠٠٠
NUM: CZTERY-TYSIĄCE 1.1 P:4;L:Ø	1	ك١٠٠٠
NUM: CZTERY-TYSIĄCE 1.2 P:4;L:4	1	ك١٠٠٠
NUM: CZTERYSTA P:4;L:Ø	21	
NUM: CZTERYSTA © P:4;L:Ø	17	
NUM: CZWARTE-PIĘTRO 1.1 P:4;L:Ø	1	ك١٠٠٠
NUM: DRUGA-GODZINA © 1.1 P:V;L:Ø	6	ك١٠٠٠
NUM: DRUGA-GODZINA © 1.2 P:L;L:Ø	17	ك١٠٠٠
NUM: DRUGA-GRUPA-INWALIDZKA 1.1 P:H;L:B	1	ك١٠٠٠
NUM: DRUGI 1.1 P:V;L:Ø (OBRÓT)	256	ك١٠٠٠
NUM: DRUGI 1.2 P:V;L:Ø (BEZ OBRÓT)	28	ك١٠٠٠
NUM: DRUGI 1.3 P:V;L:Ø/P:N;L:Ø (OBRÓT) (BOK)	14	ك١٠٠٠
NUM: DRUGI 1.4 P:V;L:V (OBRÓT)	1	ك١٠٠٠
NUM: DRUGI 1.5 P:V;L:Z (OBRÓT) (DWA)	9	ك١٠٠٠
NUM: DRUGI 1.6 P:V;L:B (OBRÓT) (DWA)	3	ك١٠٠٠
NUM: DRUGI 1.7 P:V;L:N (OBRÓT) (DWA)	4	ك١٠٠٠
NUM: DRUGI-PUNKT/DRUGI P:V;L:Z	0	
NUM: DRUGI/DWA 2.1 P:L;L:Ø (OBRÓT)	66	ك١٠٠٠
NUM: DRUGI/DWA 2.2 P:L;L:L (OBRÓT)	2	ك١٠٠٠
NUM: DRUGI/DWA 2.3 P:L;L:Z (OBRÓT)	1	ك١٠٠٠
NUM: DRUGI/DWA 2.4 P:L;L:Ø	30	ك١٠٠٠

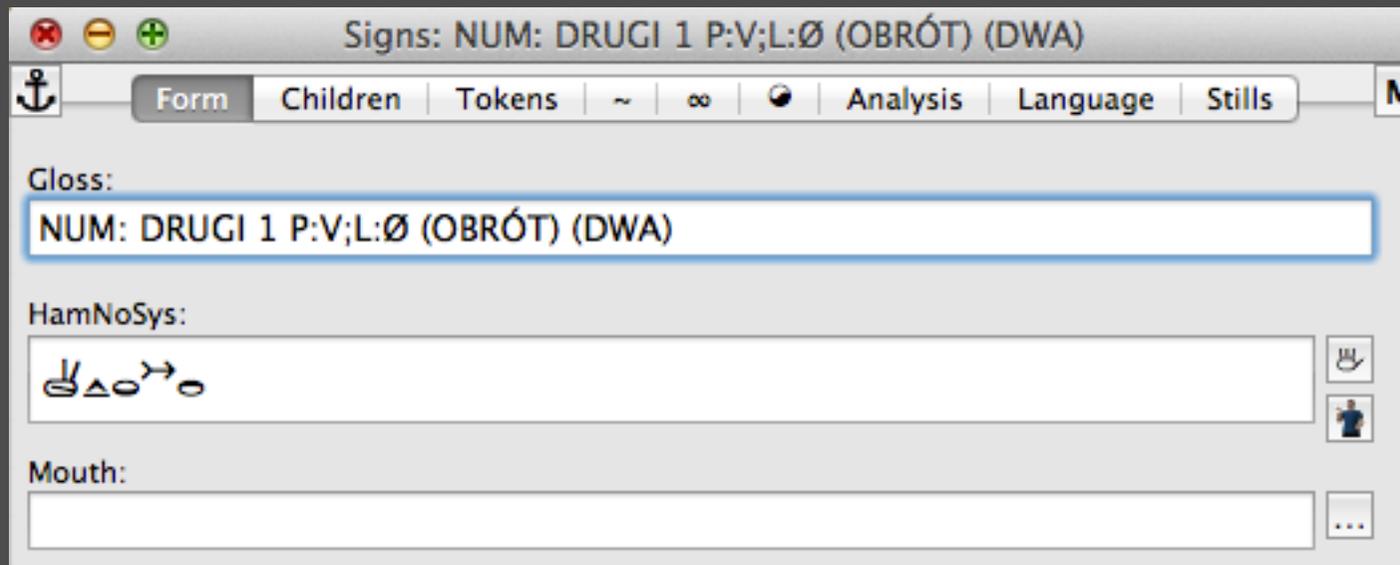
Number incorporation

- Glosses for incorporated numbers receive the prefix NUM: and are written in words. Polish spoken words used for the number and the noun.

The screenshot shows a software window titled "Signs: NUM: SZEŚĆ-MIESIĘCY 1 P:1;L:5". The window has a menu bar with "Form", "Children", "Tokens", "~", "∞", a globe icon, "Analysis", "Language", "Stills", and "M". Below the menu bar, there is a "Gloss:" field containing the text "NUM: SZEŚĆ-MIESIĘCY 1 P:1;L:5". Below the gloss field, there is a "HamNoSys:" field containing the sign language notation "· [[O, x, 3]] ΔO [< x > ΔO]". To the right of the HamNoSys field, there are two icons: a hand icon and a person icon. Below the HamNoSys field, there is a "Mouth:" field which is currently empty. To the right of the Mouth field, there is a three-dot menu icon.

Ordinal numbers

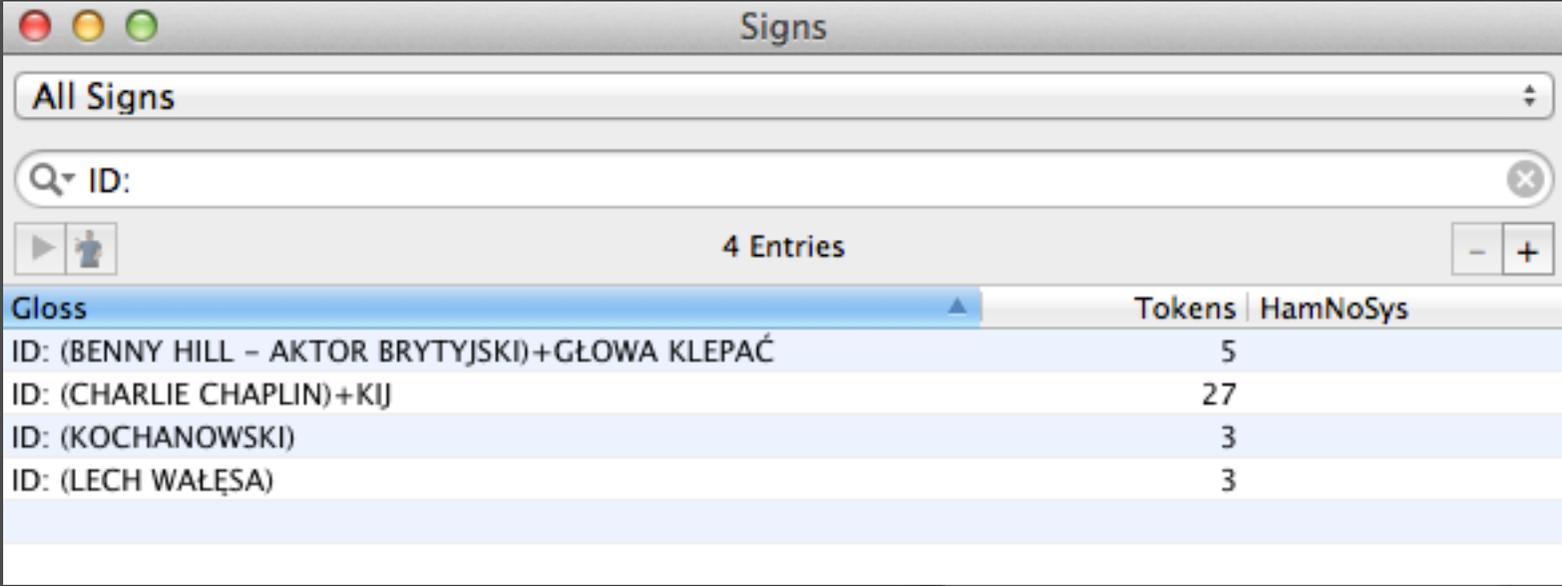
- Glosses for ordinal numbers receive the prefix NUM: and are written in words.



The screenshot shows a window titled "Signs: NUM: DRUGI 1 P:V;L:Ø (OBRÓT) (DWA)". The window has a menu bar with "Form", "Children", "Tokens", "~", "∞", a circular arrow icon, "Analysis", "Language", and "Stills". Below the menu bar, there is a "Gloss:" field containing the text "NUM: DRUGI 1 P:V;L:Ø (OBRÓT) (DWA)". Below the gloss field, there is a "HamNoSys:" field containing the text "𐄂𐄃𐄄𐄅𐄆". To the right of the HamNoSys field, there are two icons: a hand icon and a person icon. Below the HamNoSys field, there is a "Mouth:" field which is currently empty. To the right of the Mouth field, there is a three-dot menu icon.

Sign names

- ⦿ Glosses for sign names are preceded by the prefix ID: followed by the name and surname in brackets.



The screenshot shows a window titled "Signs" with a search bar containing "All Signs" and a search field with "ID:". Below the search bar, there are navigation icons and a label "4 Entries". The main content is a table with three columns: "Gloss", "Tokens", and "HamNoSys".

Gloss	Tokens	HamNoSys
ID: (BENNY HILL - AKTOR BRYTYJSKI)+GŁOWA KLEPAĆ	5	
ID: (CHARLIE CHAPLIN)+KIJ	27	
ID: (KOCHANOWSKI)	3	
ID: (LECH WAŁĘSA)	3	

Fingerspelling

- All occurrences of fingerspelling are divided into subtypes within the type **FINGERSPELLING**. Letters in words are divided with dots.

Level	Gloss	HamNoSys
Lexemes	U.R.A.Z.Y.	U.R.A.Z.Y.
Lexemes	N.A.B.I.A.Ł.	N.A.B.I.A.Ł.
Lexemes	C.O.	C.O.
Lexemes	R.	R.
Lexemes	A.B.C.	A.B.C.
Lexemes	A.CH.	A.CH.
Lexemes	Ż.Y.R.A.D.Ó.W.	Ż.Y.R.A.D.Ó.W.
Lexemes	Ś.L.U.Z.A.	Ś.L.U.Z.A.
Lexemes	A.N.D.R.Z.E.J.	A.N.D.R.Z.E.J.
Lexemes	A.D.O.B.E.	A.D.O.B.E.
Lexemes	A.G.N.I.E.S.Z.K.A.	A.G.N.I.E.S.Z.K.A.
Lexemes	A.K.	A.K.
Lexemes	A.L.A.R.M.	A.L.A.R.M.
Lexemes	A.L.G.I.E.R.I.A.	A.L.G.I.E.R.I.A.
Lexemes	A.M.A.T.O.R.	A.M.A.T.O.R.
Lexemes	A.Ż.	A.Ż.
Lexemes	A.R.G.E.N.T.Y.N.A.	A.R.G.E.N.T.Y.N.A.
Lexemes	A.R.B.U.Z.	A.R.B.U.Z.
Lexemes	A.N.N.A.	A.N.N.A.
Lexemes	A.N.I.A.	A.N.I.A.
Lexemes	A.N.G.L.I.A	A.N.G.L.I.A
Lexemes	A.N.E.T.A.	A.N.E.T.A.
Lexemes	B.	B.
Lexemes	B.A.L.	B.A.L.
Lexemes	B.A.L.E.T.	B.A.L.E.T.
Lexemes	B.A.R.	B.A.R.
Lexemes	B.A.Ł.T.Ó.W.	B.A.Ł.T.Ó.W.
Lexemes	B.E.N.N.Y. H.I.L.L.	B.E.N.N.Y. H.I.L.L.
Lexemes	B.H.P.	B.H.P.
Lexemes	B.I.A.Ł.O.R.U.Ś.	B.I.A.Ł.O.R.U.Ś.
Lexemes	B.O.+B.O. P:BO;L:BO	B.O.+B.O. P:BO;L:BO
Lexemes	B.O.X.	B.O.X.
Lexemes	B.P.	B.P.

Pointing signs

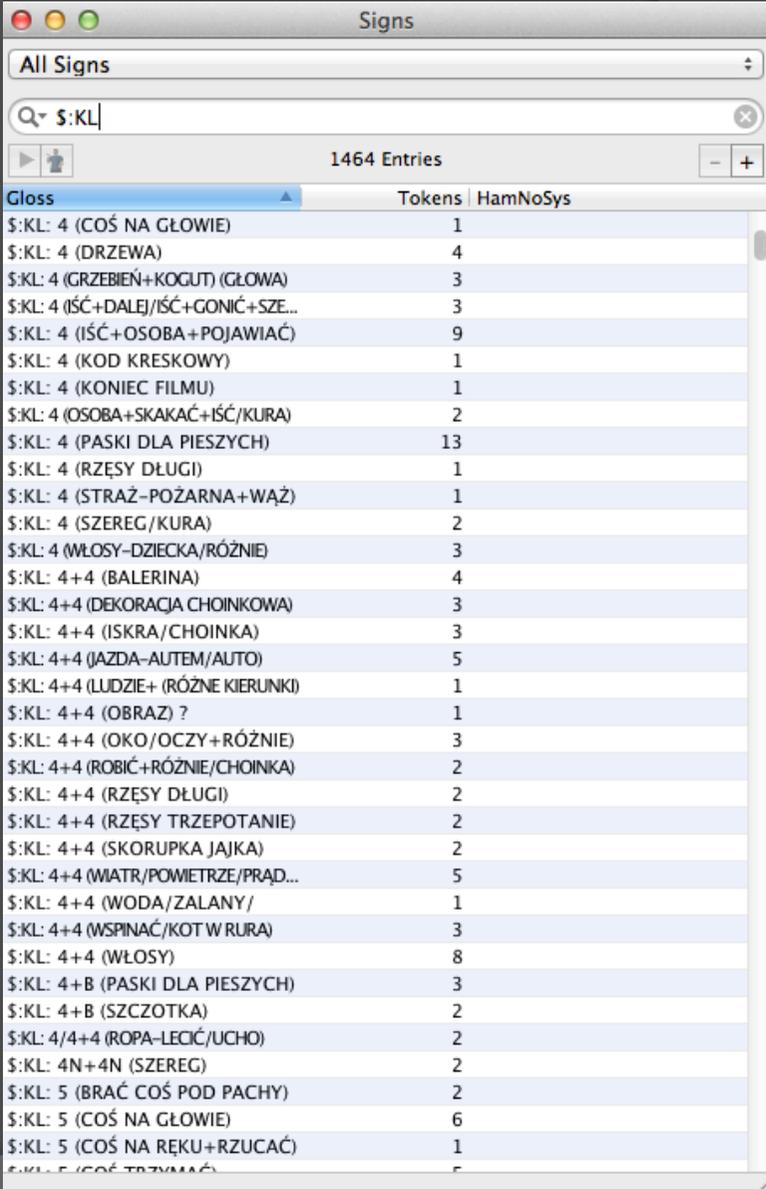
- Glosses for pointing signs are preceded with the prefix **WSKAZ:**
- Articulation info: hand configuration and localization.

The screenshot shows a window titled "Signs" with a search bar containing "wskaz" and a list of 65 entries. The list is organized into three columns: Gloss, Tokens, and HamNoSys. The entries are sorted by token count in descending order.

Gloss	Tokens	HamNoSys
WSKAZ/W-ŚRODKU P:G L:Ø (CENTR...	171	X
WSKAZ: 1 (JA)	40	
WSKAZ: 1 (RÓŻNE KIERUNKI)	97	
WSKAZ: 1+1 (DLA SIEBIE)	2	
WSKAZ: 1+1 (Z-TYŁU)	1	X
WSKAZ: 5 (WY/ONI)	5	
WSKAZ: 5 1 (JA/MÓJ)	0	⌘ ∞ 0 ☐ • X
WSKAZ: 5+5 (JA)	4	
WSKAZ: B (JA)	1670	
WSKAZ: B (KIERUNEK LUB MAPA/JE...	320	
WSKAZ: B (NOGA)	18	
WSKAZ: B (O TOBIE/O CIEBIE)	4	
WSKAZ: B (PIONOWO)	8	
WSKAZ: B (RUCH PO OKRĘGU W PION...	2	
WSKAZ: B (RUCH PO OKRĘGU W POZI...	26	
WSKAZ: B (RÓŻNE KIERUNKI)	240	
WSKAZ: B (SERCE)	9	
WSKAZ: B (SKOŚNY)	4	
WSKAZ: B+B (JA)	57	
WSKAZ: B+B (JA/TOŻSAMOŚĆ/O MN...	10	
WSKAZ: B+B (KIERUNEK LUB MAPA/J...	55	
WSKAZ: B+B (MY) (JA CZY TY)	2	
WSKAZ: B+B (Z-TYŁU)	1	
WSKAZ: C (CIEBIE/OSOBA)/ INDYWID...	164	
WSKAZ: C (JA) (MNIE)	67	
WSKAZ: C (WSZYSCY+SZEREG)	1	
WSKAZ: N (JA)(NIE)	39	
WSKAZ: N (RUCH PO OKRĘGU W PIO...	5	
WSKAZ: Z (BRODA)/SMAK/	18	
WSKAZ: Z (BRZUCH)	8	X
WSKAZ: Z (GŁOWA)/PRZYPOMNIEĆ/...	309	
WSKAZ: Z (JA)	6548	
WSKAZ: Z (JA+DO PRZODU) – DO PO...	106	
WSKAZ: Z (KIERUNEK LUB MAPA)	257	
WSKAZ: Z (KRZYŻ)	2	
WSKAZ: Z (NOGA) (KOLANO) (CIEŁO)	10	

Classifier constructions & depicting signs

- Glosses for classifiers have the prefix \$:KL:, followed by the hand configuration and an approximation of meaning given in brackets.



Gloss	Tokens	HamNoSys
\$:KL: 4 (COŚ NA GŁOWIE)	1	
\$:KL: 4 (DRZEWA)	4	
\$:KL: 4 (GRZEBIEN+KOGUT) (GŁOWA)	3	
\$:KL: 4 (IŚĆ+DALEJ)/IŚĆ+GONIĆ+SZE...	3	
\$:KL: 4 (IŚĆ+OSOBA+POJAWIAĆ)	9	
\$:KL: 4 (KOD KRESKOWY)	1	
\$:KL: 4 (KONIEC FILMU)	1	
\$:KL: 4 (OSOBA+SKAKAĆ+IŚĆ/KURA)	2	
\$:KL: 4 (PASKI DLA PIESZYCH)	13	
\$:KL: 4 (RZĘSY DŁUGI)	1	
\$:KL: 4 (STRAŻ-POŻARNA+WĄŻ)	1	
\$:KL: 4 (SZEREG/KURA)	2	
\$:KL: 4 (WŁOSY-DZIECKA/RÓŻNIE)	3	
\$:KL: 4+4 (BALERINA)	4	
\$:KL: 4+4 (DEKORACJA CHOINKOWA)	3	
\$:KL: 4+4 (ISKRA/CHOINKA)	3	
\$:KL: 4+4 (JAZDA-AUTEM/AUTO)	5	
\$:KL: 4+4 (LUDZIE+ (RÓŻNE KIERUNKI)	1	
\$:KL: 4+4 (OBRAZ) ?	1	
\$:KL: 4+4 (OKO/OCZY+RÓŻNIE)	3	
\$:KL: 4+4 (ROBIĆ+RÓŻNIE/CHOINKA)	2	
\$:KL: 4+4 (RZĘSY DŁUGI)	2	
\$:KL: 4+4 (RZĘSY TRZEPOTANIE)	2	
\$:KL: 4+4 (SKORUPKA JAJKA)	2	
\$:KL: 4+4 (WIATR/POMIETRZE/PRĄD...	5	
\$:KL: 4+4 (WODA/ZALANY/	1	
\$:KL: 4+4 (WSPINAĆ/KOT W RURA)	3	
\$:KL: 4+4 (WŁOSY)	8	
\$:KL: 4+B (PASKI DLA PIESZYCH)	3	
\$:KL: 4+B (SZCZOTKA)	2	
\$:KL: 4/4+4 (ROPA-LECIĆ/UCHO)	2	
\$:KL: 4N+4N (SZEREG)	2	
\$:KL: 5 (BRAĆ COŚ POD PACHY)	2	
\$:KL: 5 (COŚ NA GŁOWIE)	6	
\$:KL: 5 (COŚ NA RĘKU+RZUCAĆ)	1	
\$:KL: 5 (COŚ TRZYMAĆ)	1	

Classifier constructions & depicting signs

Signs: \$:KL: N+N (IŚĆ/BIEGAĆ STOPAMI/ZWIERZĘTA MAJĄCE 4 ŁAPKI)

Form Children Tokens ~ ∞ Analysis Language Stills M

Gloss:
\$:KL: N+N (IŚĆ/BIEGAĆ STOPAMI/ZWIERZĘTA MAJĄCE 4 ŁAPKI)

HamNoSys:
[] [Hand icon] [Person icon]

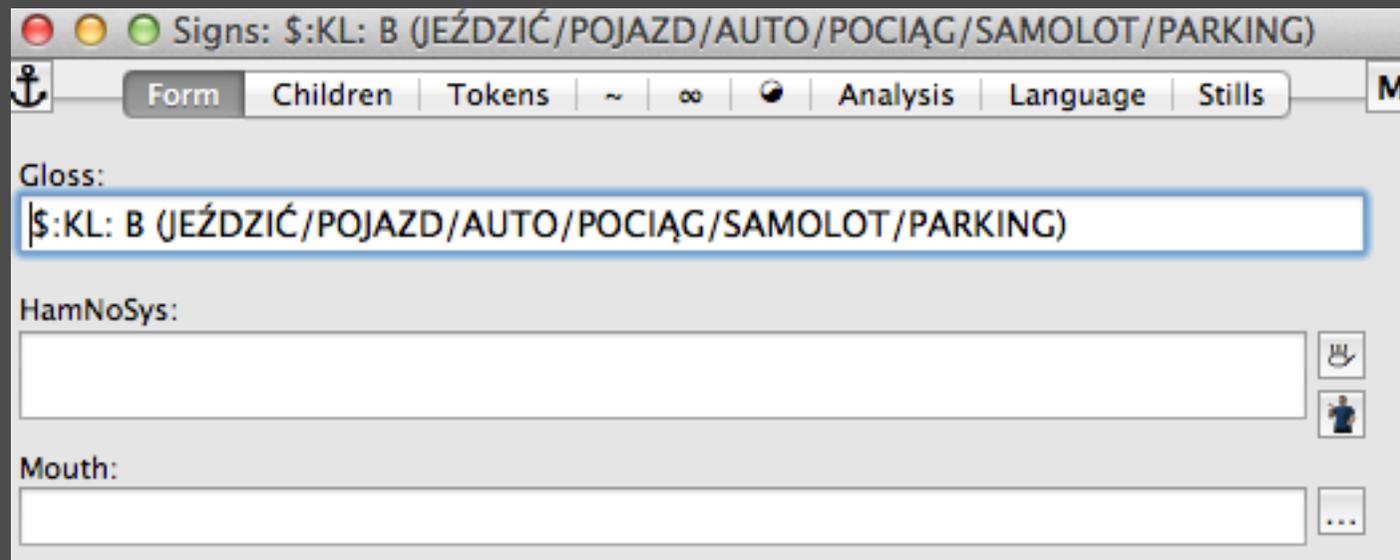
Mouth:
[] [More icon]

Composed of:
[] [More icon] [] [More icon]

\$:KL: N+N (GO/RUN ON FEET/ANIMALS WITH 4 LEGS)

Shape constructions and type-like classifier/depicting signs

- Both types are annotated with the “\$:KL” prefix.

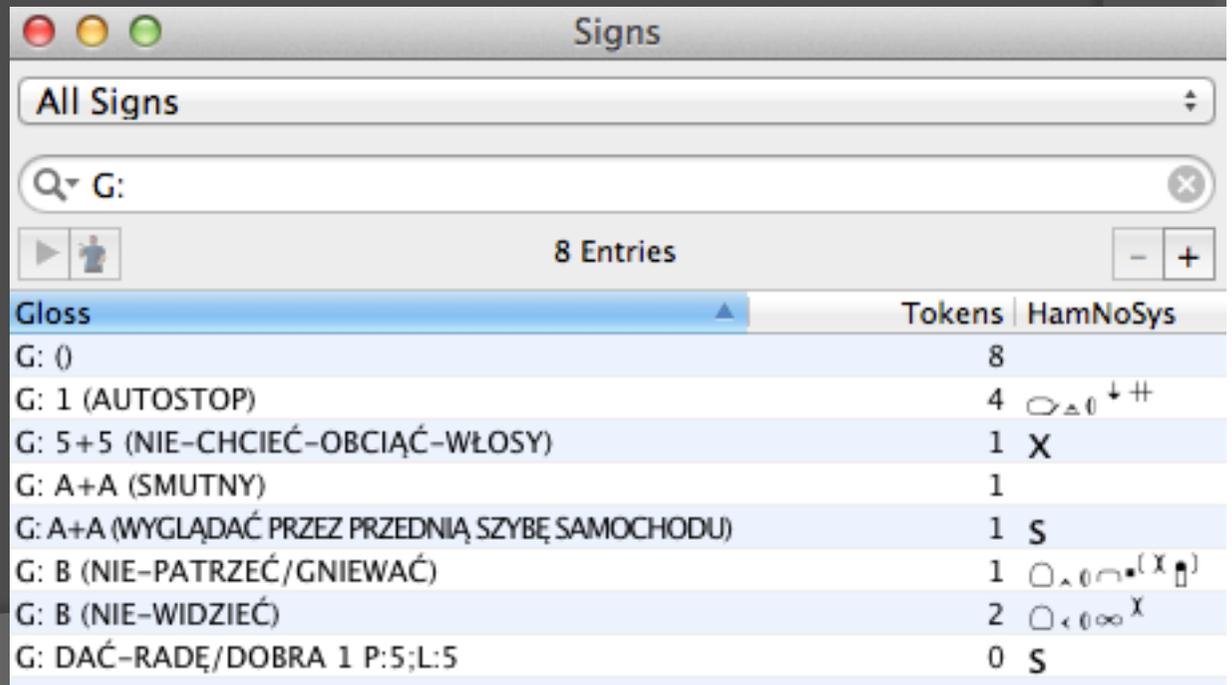


The screenshot shows a software window titled "Signs: \$:KL: B (JEŹDZIĆ/POJAZD/AUTO/POCIĄG/SAMOLOT/PARKING)". The window has a menu bar with options: Form, Children, Tokens, ~, ∞, a circular icon, Analysis, Language, and Stills. Below the menu bar, there are three input fields:

- Gloss:** A text box containing the string "\$:KL: B (JEŹDZIĆ/POJAZD/AUTO/POCIĄG/SAMOLOT/PARKING)".
- HamNoSys:** An empty text box with a small icon of a hand on the right side.
- Mouth:** An empty text box with a small icon of three dots on the right side.

Gestures

- Glosses for gestures received the prefix G: followed by the handshape and an approximation of meaning given in brackets. Phatic gestures annotated separately: @ - dragging attention; & - never mind and ^ - pauses.

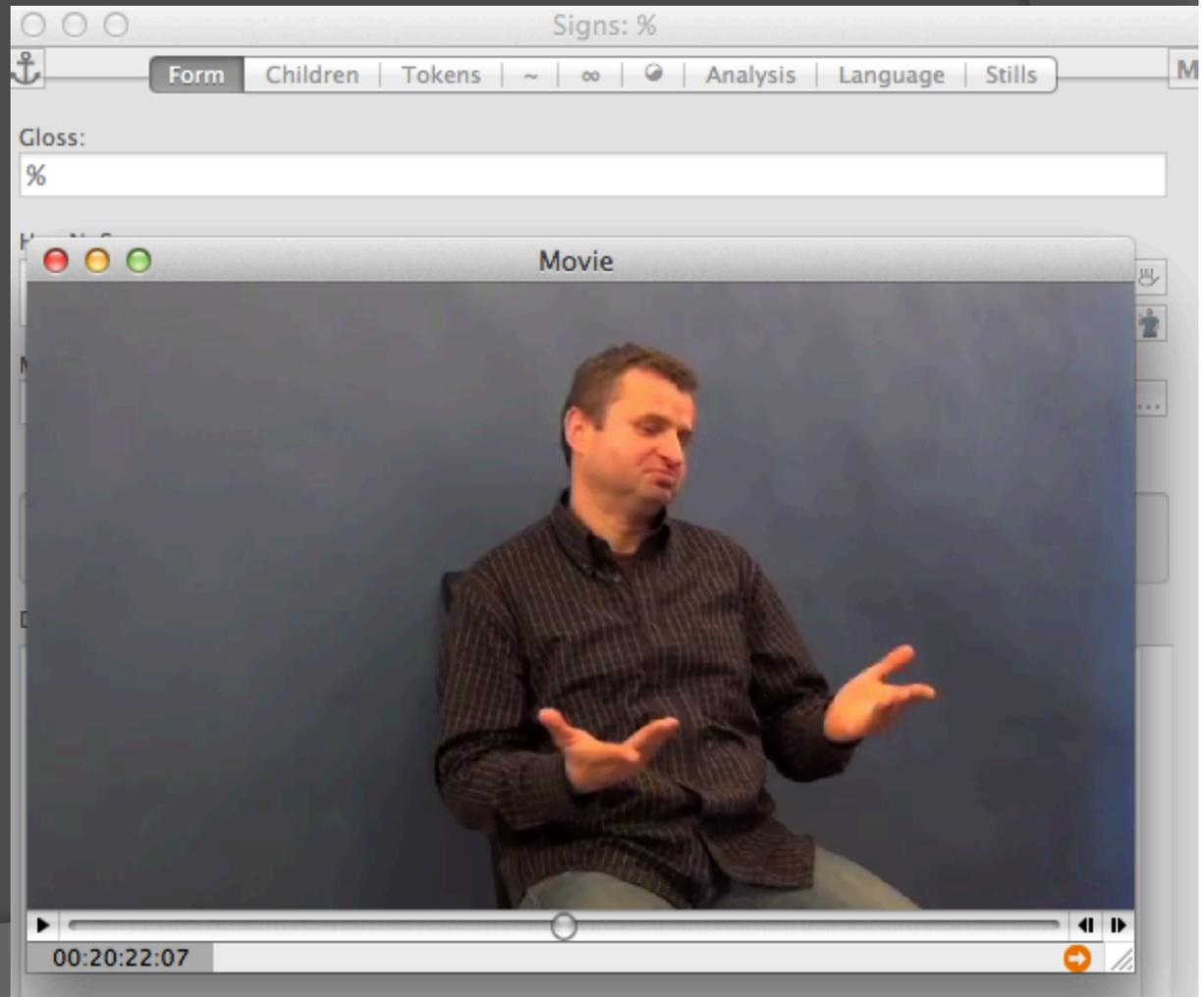


The screenshot shows a window titled "Signs" with a search bar containing "G:". Below the search bar, it indicates "8 Entries". The main content is a table with the following columns: Gloss, Tokens, and HamNoSys.

Gloss	Tokens	HamNoSys
G: ()	8	
G: 1 (AUTOSTOP)	4	☉△0 ↓++
G: 5+5 (NIE-CHCIEĆ-OBCIĄĆ-WŁOSY)	1	X
G: A+A (SMUTNY)	1	
G: A+A (WYGLĄDAĆ PRZEZ PRZEDNIĄ SZYBĘ SAMOCHODU)	1	S
G: B (NIE-PATRZEĆ/GNIEWAĆ)	1	☉△0☉*(X☉)
G: B (NIE-WIDZIEĆ)	2	☉△0∞X
G: DAĆ-RADE/DOBRA 1 P:5;L:5	0	S

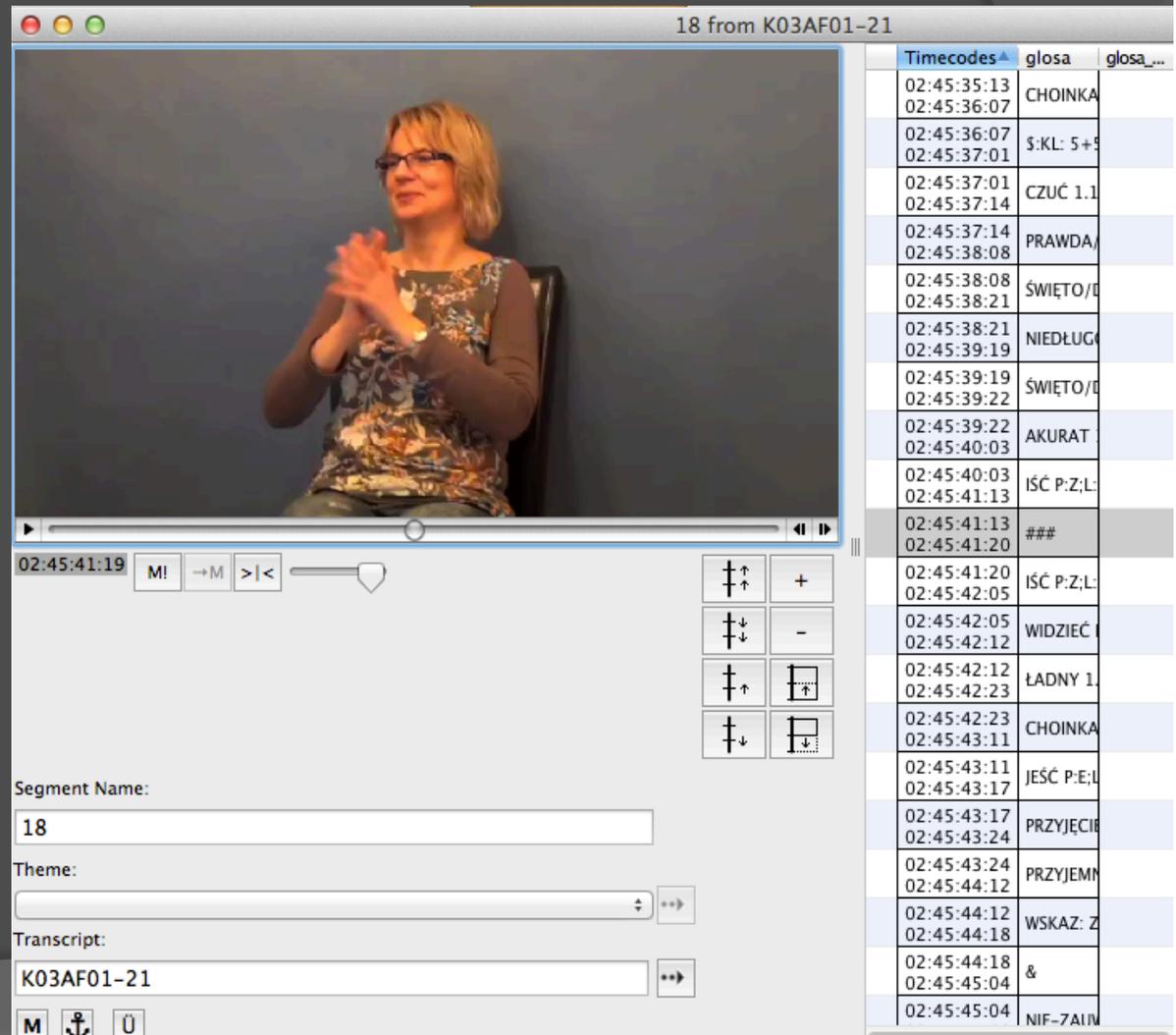
Palm-up

- Annotated with the % symbol.



Doubt as to whether the movement is a sign or not

- Indecipherable, invisible, unclear or doubtful signs are annotated with the ### symbol.

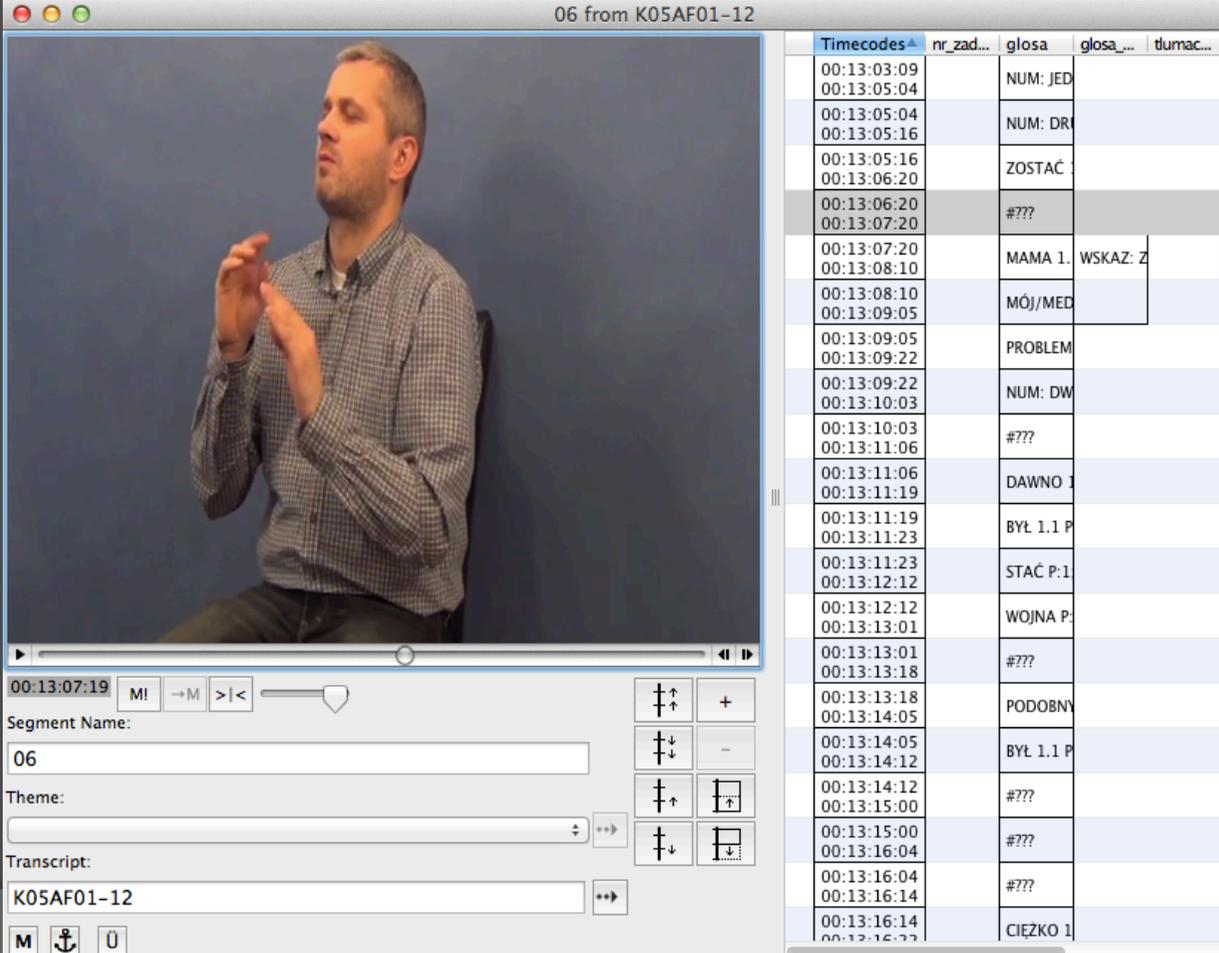


The screenshot shows a video player window titled "18 from K03AF01-21". The video frame displays a woman with blonde hair and glasses, wearing a patterned top, clapping her hands. Below the video frame is a control bar with a play button, a progress slider, and a volume icon. The current time is 02:45:41:19. To the right of the video frame is a list of timecodes and their corresponding annotations in Polish. The annotations include words like "CHOINKA", "ŚWIĘTO", "NIEDŁUGO", "AKURAT", "IŚĆ P:Z:L:", "WIDZIEĆ", "ŁADNY 1.", "PRZYJĘCIE", "PRZYJEMN", "WSKAZ: Z", and "NIF-7ALIV". The annotation "###" is highlighted in grey, indicating a doubtful sign.

Timecodes	glosa	glosa_...
02:45:35:13 02:45:36:07	CHOINKA	
02:45:36:07 02:45:37:01	Ś:KL: 5+5	
02:45:37:01 02:45:37:14	CZUĆ 1.1	
02:45:37:14 02:45:38:08	PRAWDA/	
02:45:38:08 02:45:38:21	ŚWIĘTO/	
02:45:38:21 02:45:39:19	NIEDŁUGO	
02:45:39:19 02:45:39:22	ŚWIĘTO/	
02:45:39:22 02:45:40:03	AKURAT	
02:45:40:03 02:45:41:13	IŚĆ P:Z:L:	
02:45:41:13 02:45:41:20	###	
02:45:41:20 02:45:42:05	IŚĆ P:Z:L:	
02:45:42:05 02:45:42:12	WIDZIEĆ	
02:45:42:12 02:45:42:23	ŁADNY 1.	
02:45:42:23 02:45:43:11	CHOINKA	
02:45:43:11 02:45:43:17	JEŚĆ P:E:L	
02:45:43:17 02:45:43:24	PRZYJĘCIE	
02:45:43:24 02:45:44:12	PRZYJEMN	
02:45:44:12 02:45:44:18	WSKAZ: Z	
02:45:44:18 02:45:45:04	&	
02:45:45:04	NIF-7ALIV	

New gloss needed, no proposal yet

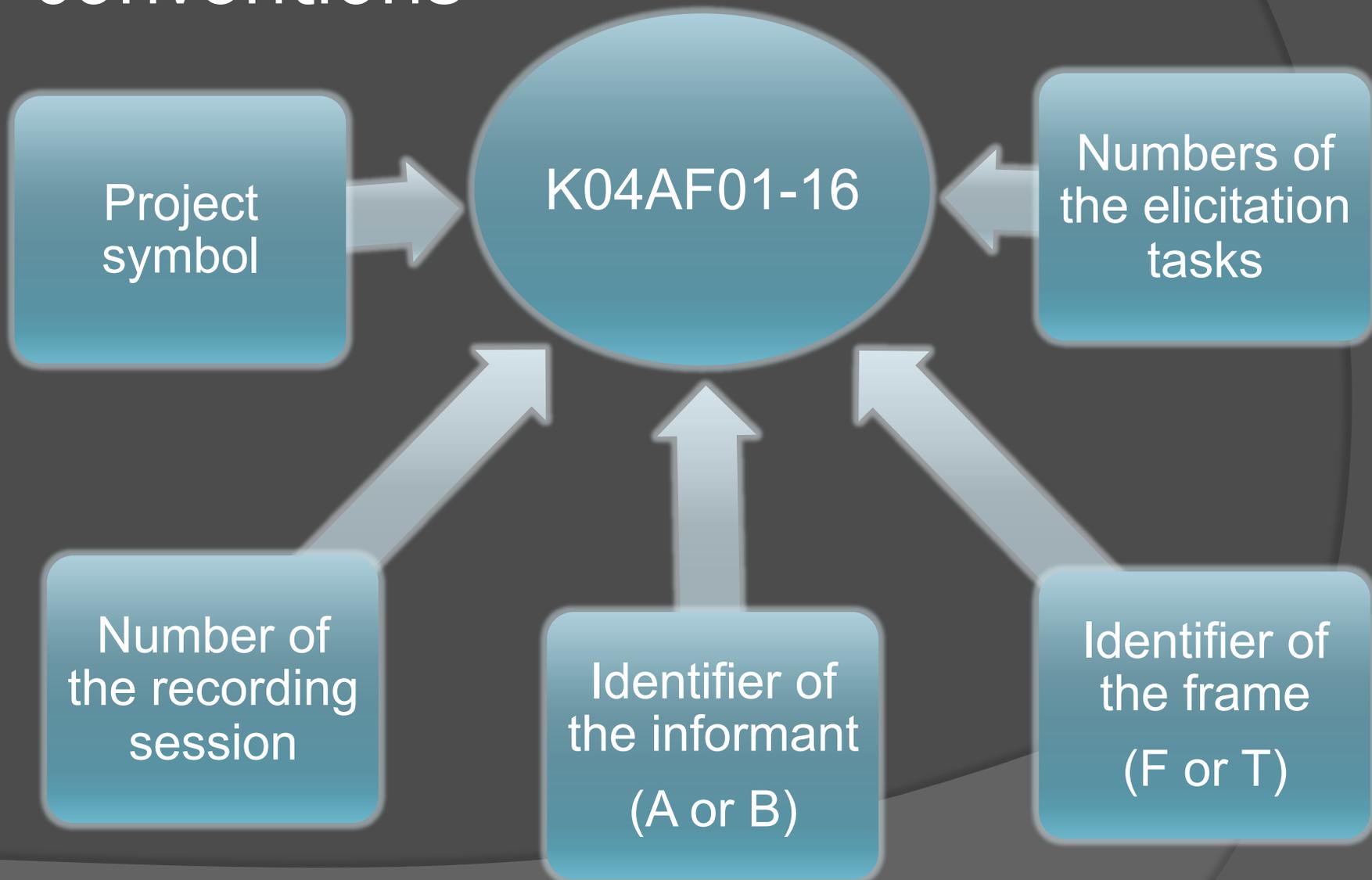
- Annotated with the #??? symbol.



The screenshot displays a video player window titled "06 from K05AF01-12". The video shows a man in a checkered shirt speaking. Below the video is a control bar with a timestamp of 00:13:07:19 and playback controls. To the right is a subtitle table with columns: Timecodes, nr_zad..., glosa, glosa_..., and tłumac... The table contains several rows of subtitles, with one row highlighted in grey containing the text "#???" between the timecodes 00:13:06:20 and 00:13:07:20.

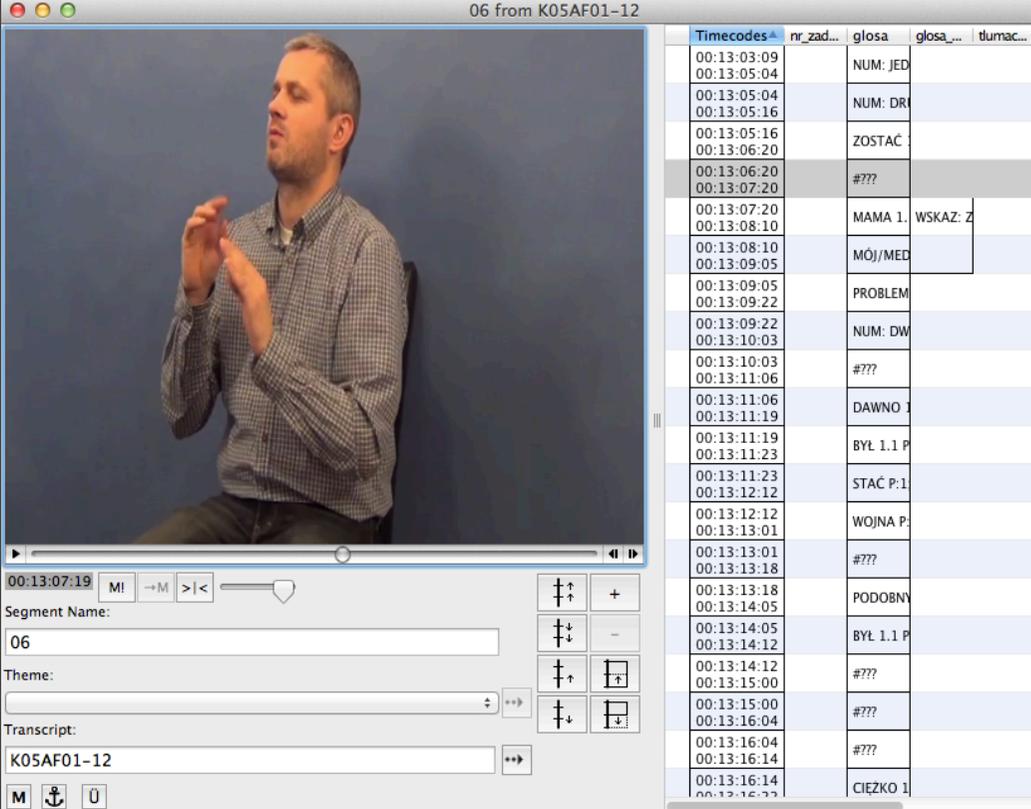
Timecodes	nr_zad...	glosa	glosa_...	tłumac...
00:13:03:09 00:13:05:04		NUM: JED		
00:13:05:04 00:13:05:16		NUM: DR		
00:13:05:16 00:13:06:20		ZOSTAĆ :		
00:13:06:20 00:13:07:20		#???		
00:13:07:20 00:13:08:10		MAMA 1. WSKAZ: Z		
00:13:08:10 00:13:09:05		MÓJ/MED		
00:13:09:05 00:13:09:22		PROBLEM		
00:13:09:22 00:13:10:03		NUM: DW		
00:13:10:03 00:13:11:06		#???		
00:13:11:06 00:13:11:19		DAWNO I		
00:13:11:19 00:13:11:23		BYŁ 1.1 P		
00:13:11:23 00:13:12:12		STAĆ P:1		
00:13:12:12 00:13:13:01		WOJNA P:		
00:13:13:01 00:13:13:18		#???		
00:13:13:18 00:13:14:05		PODOBNY		
00:13:14:05 00:13:14:12		BYŁ 1.1 P		
00:13:14:12 00:13:15:00		#???		
00:13:15:00 00:13:16:04		#???		
00:13:16:04 00:13:16:14		#???		
00:13:16:14 00:13:16:22		CIEŻKO 1		

Annotation and video file naming conventions



Video file viewing conventions

- Two annotation files per signer (eg. K04AF01-16 and K04AF17-21) linked to two movie files.
- One informant visible in the annotating window.
- General view is possible.



The screenshot displays a video annotation software interface. On the left, a video player shows a man speaking. Below the player, a control bar includes a timecode display (00:13:07:19), playback controls, and a segment name field containing '06'. The 'Transcript' field shows 'K05AF01-12'. On the right, a table lists timecodes and associated text.

Timecodes	nr_zad...	glosa	glosa_...	tlumac...
00:13:03:09 00:13:05:04		NUM: JED		
00:13:05:04 00:13:05:16		NUM: DR		
00:13:05:16 00:13:06:20		ZOSTAĆ		
00:13:06:20 00:13:07:20		#???		
00:13:07:20 00:13:08:10		MAMA 1.	WSKAZ. Z	
00:13:08:10 00:13:09:05		MÓJ/MED		
00:13:09:05 00:13:09:22		PROBLEM		
00:13:09:22 00:13:10:03		NUM: DW		
00:13:10:03 00:13:11:06		#???		
00:13:11:06 00:13:11:19		DAWNO 1		
00:13:11:19 00:13:11:23		BYŁ. 1.1 P		
00:13:11:23 00:13:12:12		STAC P.1		
00:13:12:12 00:13:13:01		WOJNA P		
00:13:13:01 00:13:13:18		#???		
00:13:13:18 00:13:14:05		PODOBNY		
00:13:14:05 00:13:14:12		BYŁ. 1.1 P		
00:13:14:12 00:13:15:00		#???		
00:13:15:00 00:13:16:04		#???		
00:13:16:04 00:13:16:14		#???		
00:13:16:14 00:13:16:22		CIEŻKO 1		

Feedback

- Annotating two-handed signs on both dominant and non-dominant hand is redundant in our project since the articulation is encoded in the name of the gloss.
- Additional information found in the HamNoSys transcription (which is not coded on the annotation level).

Feedback

- The division of glosses for buoys and pointing signs proposed in the “general practise” is too detailed and requires advanced competence in linguistic analysis from the annotator.
- Dividing all signs into two large categories (lexicalised signs and other types of signs: pointing signs, classifiers, fingerspelling, gestures, palm-ups...) is enough on the basic level of annotation. At this stage we do not divide classifiers/depicting signs into finer classes.

Discussion

- ① Who is able to produce better annotation?
 - A native Deaf signer without any linguistic training?
 - A bilingual Deaf person with some linguistic background? (Possible influence of spoken language on analysed data.)
- ② How to fill in the gap between scientific needs that a corpus should serve and the daily annotation practice?
- ③ Should annotations be double-checked? If so, by whom?

THANK YOU!

Contact us:

www.plm.uw.edu.pl

p.rutkowski@uw.edu.pl

j.filipczak@uw.edu.pl

anna.kuder@uw.edu.pl