

The SignBeach Dataset of Dutch Sign Language (NGT) signs



Annika
Schiefner



Gomèr
Otterspeer



Beyza
Sümer



Floris
Roelofsen

Project development

Project idea

- Game for teaching primary school children NGT
- Gather data to investigate sign learning over time

Item design

- Target signs from dictionary and NGT Signbank
- Signer variation
- Orientation variation

Alignment with NGT signbank:

- glossID and Signbank URLs
- Interoperability with other NGT resources



Starting point:

- *Mijn eerste 1500 gebaren* (My first 1500 signs)

Recording sessions:

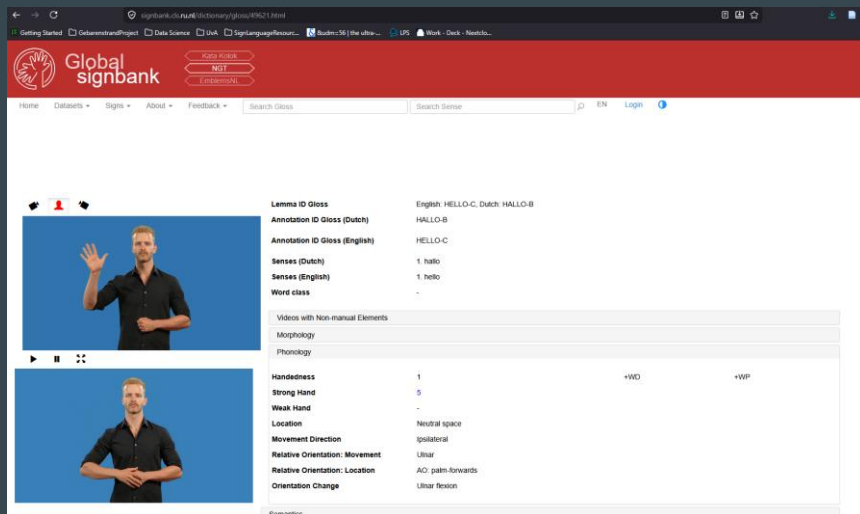
- Signer input (recording multiple variants)
- Prompting through Dutch target translations and NGT Signbank recordings

Sign beach



Related resources

Available NGT resources



The screenshot shows the Global Signbank website interface. On the left, there are two video thumbnails of a man in a black shirt making a hand gesture. The main content area displays a metadata table for the sign 'HALLO-B'.


Lemma ID Gloss	English: HELLO-C, Dutch: HALLO-B
Annotation ID Gloss (Dutch)	HALLO-B
Annotation ID Gloss (English)	HELLO-C
Senses (Dutch)	1. hallo
Senses (English)	1. hello
Word class	-

Below the table, there are sections for 'Videos with Non-manual Elements', 'Morphology', and 'Phonology'. The 'Phonology' section contains a table with the following data:


Handedness	1	+WD	+WP
Strong Hand	5		
Weak Hand	-		
Location	Neutral space		
Movement Direction	Ipsilateral		
Relative Orientation: Movement	Ulnar		
Relative Orientation: Location	AO: palm-forwards		
Orientation Change	Ulnar flexion		

- NGT Corpus
- NGT Signbank
- signCollect platform for recording





▶ || 🔍



Lemma ID Gloss	English: HELLO-C, Dutch: HALLO-B
Annotation ID Gloss (Dutch)	HALLO-B
Annotation ID Gloss (English)	HELLO-C
Senses (Dutch)	1. hallo
Senses (English)	1. hello
Word class	-

Videos with Non-manual Elements

Morphology

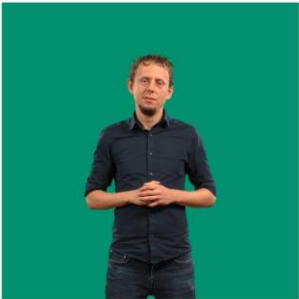
Phonology

Handedness	1
Strong Hand	5
Weak Hand	-
Location	Neutral space
Movement Direction	Ipsilateral
Relative Orientation: Movement	Ulnar
Relative Orientation: Location	AO: palm-forwards
Orientation Change	Ulnar flexion

Semantics

Iconicity and transparency data

UNIVERSITEIT VAN AMSTERDAM



Vertaling: **maandag**

Hoe sterk lijkt dit gebaar op wat het betekent?

Als je geen video of vertaling kunt zien, kies dan "technisch probleem".

1 - Niet iconisch	2	3	4	5	6	7 - Heel iconisch	Technisch probleem
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

→

- Iconicity ratings
 - Deaf NGT signers
 - Dutch hearing non-signers
 - German hearing non-signers
- Transparency scores
 - Dutch hearing non-signers
 - Translation attempts
 - Semantic similarity scores

Dataset design and methodology

Item selection and alignment

Project idea

- Game for teaching primary school children NGT
- Gather data to investigate sign learning over time

Item design

- Target signs from dictionary and NGT Signbank
- Signer variation
- Orientation variation

Alignment with NGT signbank:

- glossID and Signbank URLs
- Interoperability with other NGT resources



Starting point:

- *Mijn eerste 1500 gebaren* (My first 1500 signs)

Recording sessions:

- Signer input (recording multiple variants)
- Prompting through Dutch target translations and NGT Signbank recordings

Item selection and alignment

Project idea

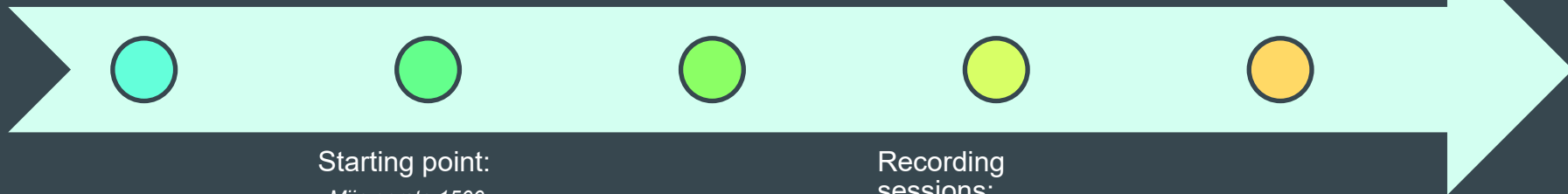
- Game for teaching primary school children NGT
- Gather data to investigate sign learning over time

Item design

- Target signs from dictionary and NGT Signbank
- Signer variation
- Orientation variation

Alignment with NGT signbank:

- glossID and Signbank URLs
- Interoperability with other NGT resources



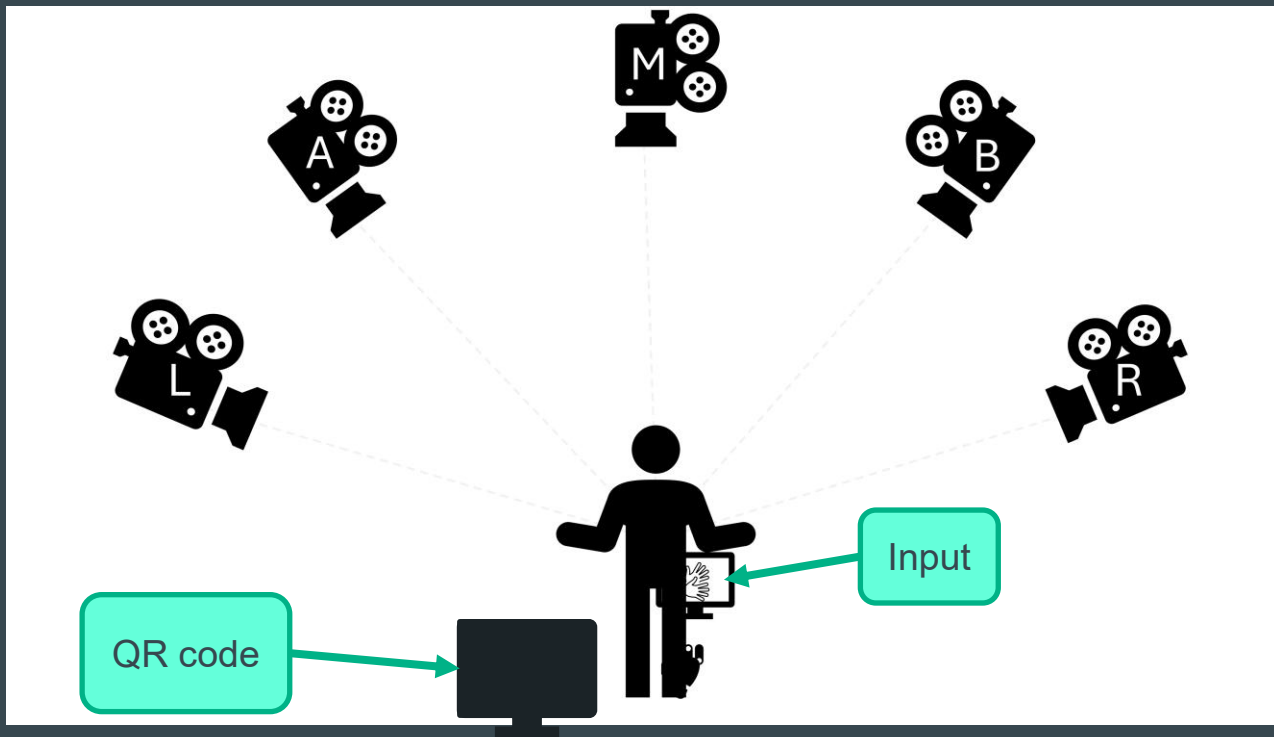
Starting point:

- *Mijn eerste 1500 gebaren* (My first 1500 signs)

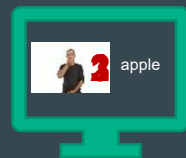
Recording sessions:

- Signer input (recording multiple variants)
- Prompting through Dutch target translations and NGT Signbank recordings

Camera setup



Multi-angle recordings



Signer

Computational derivatives

MediaPipe

- 2D/3D pose estimates
- Goal: kinematic analysis

HaMeR

- 3D hand mesh reconstructions
- Goal: handshape identification and comparisons

SAM 3D

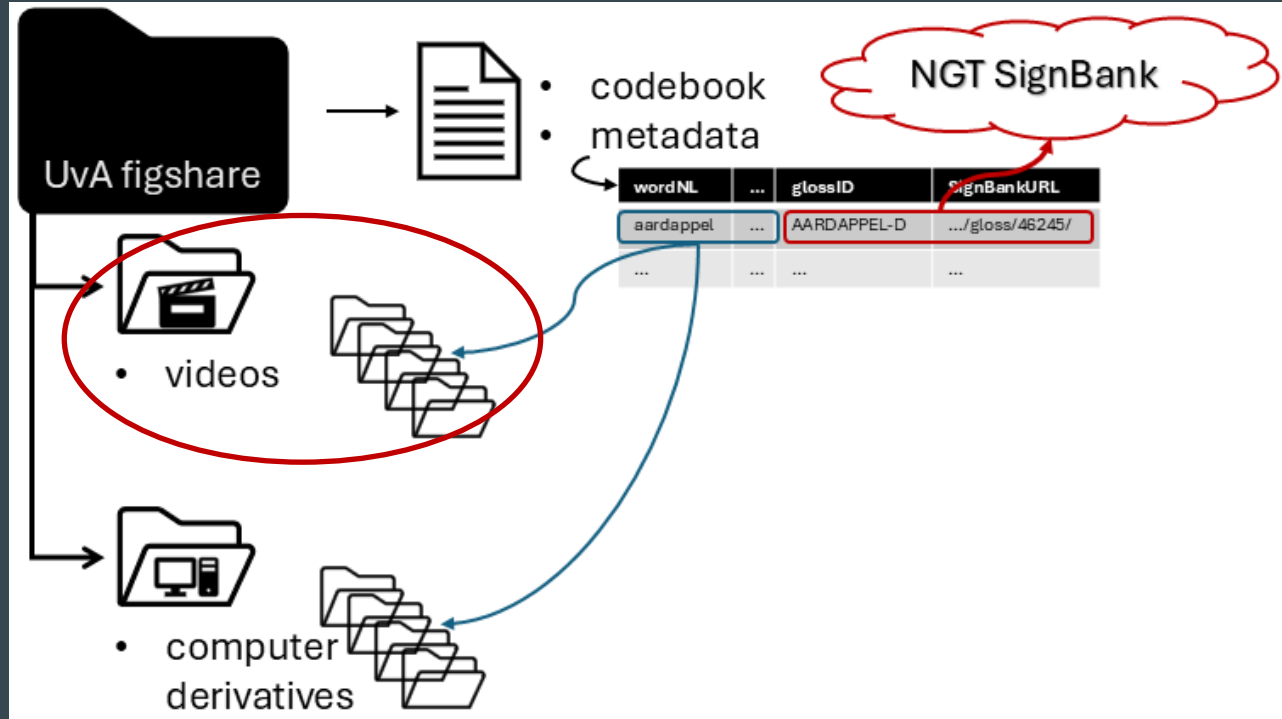
- Full-body 3D reconstructions
- Goal: avatar animation, detailed motion tracking

Dataset structure and availability

Database structure (Metadata)

Variable	Description	Possible values
wordNL	target word	Dutch words or expressions
wordEN	English translation	English words or expressions
signerID	Signer identifier	[theo, tobias, manouk, eva, debby, gomer]
orientation	Orientation of the signer relative to the central camera	[front, side]
cameraID	Camera identifier	[A, L, M, R, B]
glossID	ID gloss from the NGT Signbank	e.g., HOND, SCHOOL; compounds: GLOSS+GLOSS
SignBankURL	URL to the sign entry on the NGT Signbank	signbank.cls.ru.nl/dictionary/gloss/{id}
video	Video file	filenames*_video.mp4
mediapipe	MediaPipe Holistic keypoint estimates	filenames*_mediapipe.pose
hamer	HaMeR hand mesh keypoint estimates	filenames*_hamer.hamer
s3d	SAM 3D Body full-body reconstruction	filenames*_s3d.s3d
·filenames	Naming system	wordNL_signerID_orientation_cameraID

SignBeach Figshare structure












Videos

The screenshot shows a web interface for a video library. At the top left, the logos for the University of Amsterdam and Amsterdam University of Applied Sciences are visible. A search bar on the top right contains the text "Search on University of Am..." and a magnifying glass icon. A "Browse" link is located to the left of the search bar. On the left side, a "Files" section lists nine video files, each with a small thumbnail icon, a filename, and a size. The first file is selected. On the right side, a video player displays a man in a dark blue shirt standing against a blue background. The video player includes a progress bar at the bottom showing "00:00" and a volume icon. The text "MD5: 1c" is visible in the bottom right corner of the player area.

UNIVERSITY OF AMSTERDAM
Amsterdam University of Applied Sciences

[Browse](#) Search on University of Am...

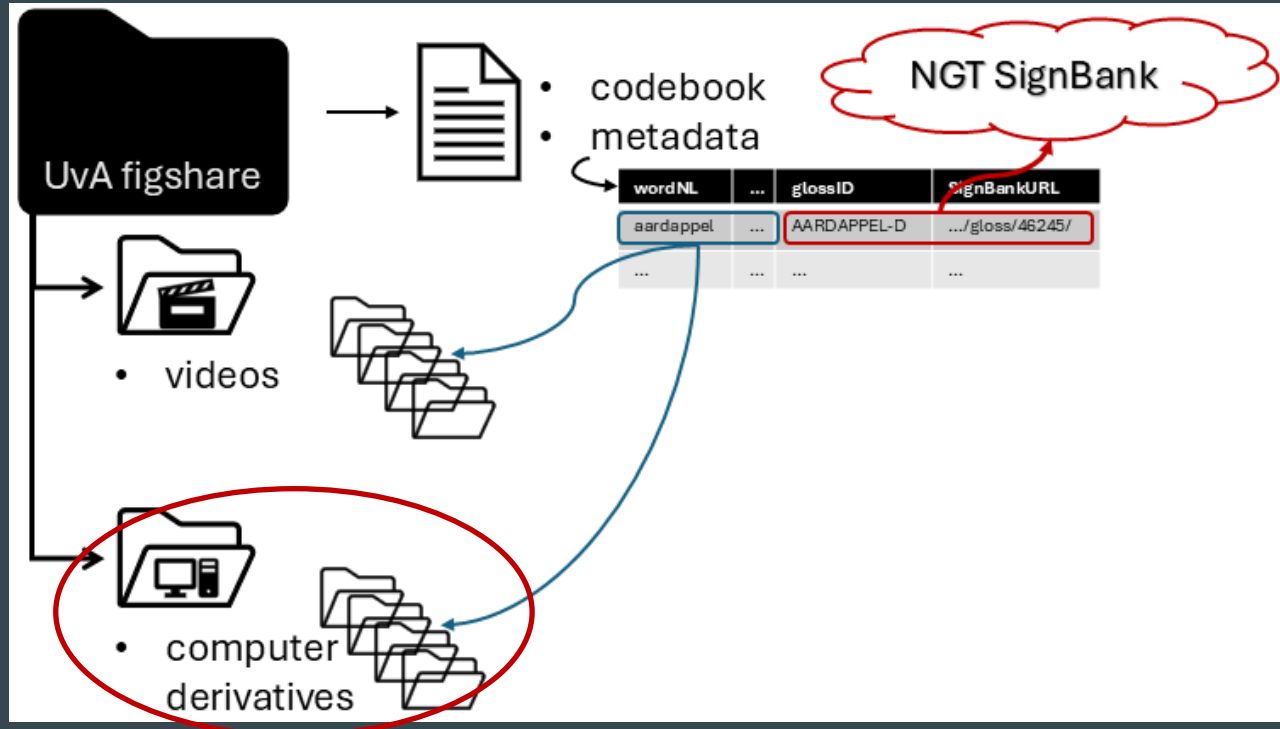
Files

-  [appelmoes_tobias_front_M_video.mp4](#) 2.39 MB ?
-  [augustus_tobias_front_R_video.mp4](#) 1.63 MB
-  [aardbei_theo_front_M_video.mp4](#) 2.25 MB
-  [ambulance_theo_front_M_video.mp4](#) 2.53 MB
-  [augustus_manouk_side_M_video.mp4](#) 1.7 MB
-  [afdrogen_tobias_front_M_video.mp4](#) 1.8 MB
-  [appelsap_tobias_front_M_video.mp4](#) 1.8 MB
-  [arm_eva_side_M_video.mp4](#) 1.47 MB
-  [amerika_tobias_side_M_video.mp4](#) 1.7 MB

00:00 00:04

MD5: 1c

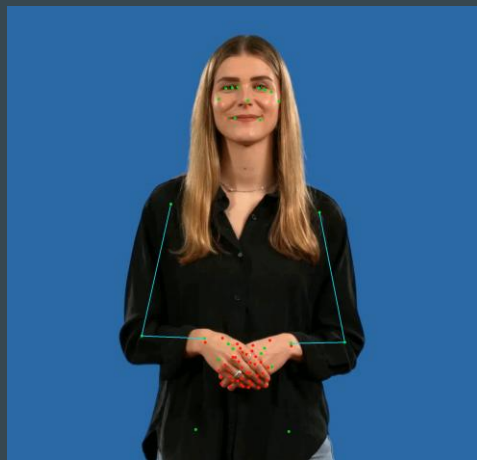
SignBeach Figshare structure



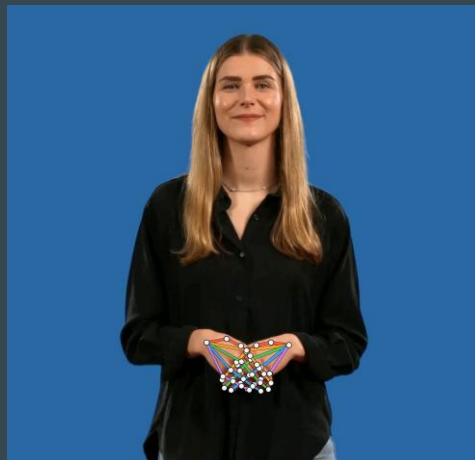
MediaPipe



HaMeR



Mediapipe



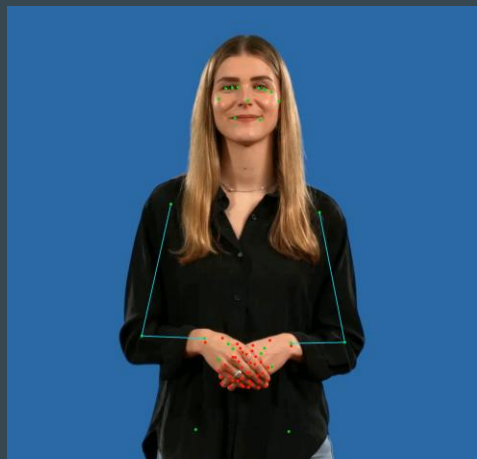
HaMeR keypoints



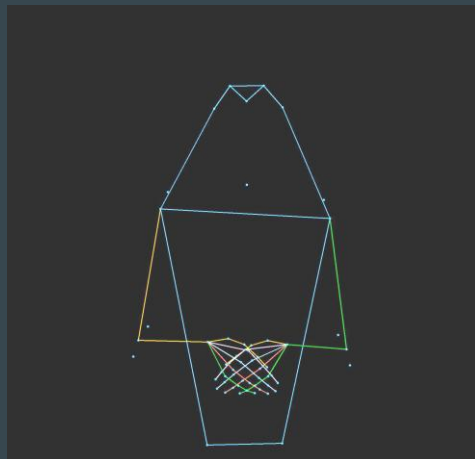
HaMeR mesh

SAM 3D Body Reconstruction

https://github.com/rem0g/sam3dbody_to_glb



Mediapipe

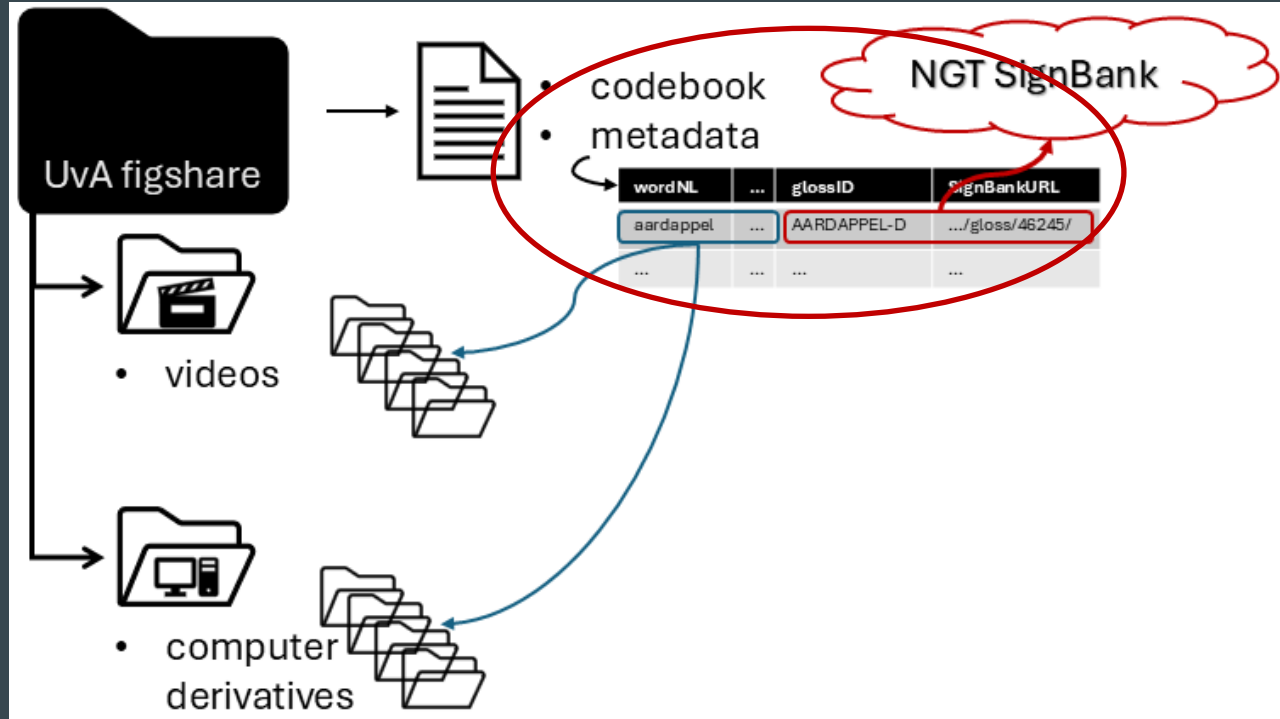


Skeleton



Body reconstruction

SignBeach Figshare structure



Limitations and future directions

Limitations

Lexical coverage: Child-friendly, everyday vocabulary

Specialised/late-learned signs are underrepresented

Recording context: Isolated sign

Studio context

Lack of diversity: Only 4 deaf signers (per sign)

No regional, generational, or sociolinguistic diversity

Future directions

Research

Factors contributing to sign learning

Alignment of perceptual and production learning

Kinematic factors in sign learning

...

Learning games

Video to picture matching

Sentence level language games

Production tasks

References

- Crasborn, O., Bank, R., Zwitserlood, I., Kooij, E. V. D., Ormel, E., Ros, J., Schüller, A., Meijer, A. D., Zuilen, M. V., Nauta, E., Winsum, F. V., & Vonk, M. (2020). NGT Signbank. <https://doi.org/10.13140/RG.2.1.2839.1446>
- Crasborn, O., Bank, R., Zwitserlood, I., van der Kooij, E., Schüller, A., Ormel, E., Nauta, E. Y., van Zuilen, M., van Winsum, F., & Ros, J. (2016). Linking Lexical and Corpus Data for Sign Languages: NGT Signbank and the Corpus NGT. In E. Efthimiou, S.-E. Fotinea, T. Hanke, J. A. Hochgesang, J. Kristoffersen, & J. Mesch (Eds), Proceedings of the LREC2016 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining (pp. 41–46). European Language Resources Association (ELRA). <https://www.sign-lang.uni-hamburg.de/lrec/pub/16023.html>
- Crasborn, O., Zwitserlood, I., Ros, J., & van Kampen, A. (2017). *Collection 'Corpus NGT' (2006-2017)* (Corpus Nos 175-346-174-413-3). The Language Archive. <https://hdl.handle.net/1839/8e5a77a3-8d1a-492a-bc86-9a3398b0809c>
- Jones, G., & Rowland, C. F. (2017). Diversity not quantity in caregiver speech: Using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth. *Cognitive Psychology*, 98, 1–21. <https://doi.org/10.1016/j.cogpsych.2017.07.002>
- Klomp, U., Gierman, L., Manders, P., Nauta, E., Otterspeer, G., Pelulessy, R., Stern, G., Venter, D., Wubbolts, C., Oomen, M., & Roelofsen, F. (2024). An extension of the NGT dataset in Global Signbank. Proceedings of the LREC-COLING 2024, 178–183. <https://aclanthology.org/2024.signlang-1.19/>
- Lugaresi, C., Shang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M., Lee, J., & others. (2019). Mediapipe: A framework for perceiving and processing reality. *Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR)*. <https://doi.org/10.48550/arXiv.1906.08172>
- Nederlands Gebarencentrum. (2012). *Mijn eerste 1500 gebaren* (1st edn). Nederlands Gebarencentrum.
- Nederlands Gebarencentrum. (2024). *Online Woordenboek NGT*. Nederlands Gebarencentrum. <https://www.gebarencentrum.nl/Gebarenwoordenboek>
- Otterspeer, G., Klomp, U., & Roelofsen, F. (2024). SignCollect: A 'touchless' pipeline for constructing large-scale sign language repositories. In E. Efthimiou, S.-E. Fotinea, T. Hanke, J. Hochgesang, J. Mesch, & M. Schulder (Eds), *Proceedings of the LREC2024 11th Workshop on the representation and processing of sign languages: Evaluation of sign language resources* (pp. 269–275). European Language Resources Association. https://www.signlab-amsterdam.nl/publications/otterspeeretal_lrec24.pdf
- Pavlakos, G., Shan, D., Radosavovic, I., Kanazawa, A., Fouhey, D., & Malik, J. (2024a). HaMeR: Hand mesh recovery for the EgoExo4D hand pose challenge. https://www.cs.utexas.edu/~pavlakos/hamer/resources/egoexo4d_challenge.pdf
- Pavlakos, G., Shan, D., Radosavovic, I., Kanazawa, A., Fouhey, D., & Malik, J. (2024b). Reconstructing hands in 3D with transformers. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 9826–9836. <https://doi.org/10.48550/arXiv.2312.05251>
- Potter, C. E., & Saffran, J. R. (2017). Exposure to multiple accents supports infants' understanding of novel accents. *Cognition*, 166, 67–72. <https://doi.org/10.1016/j.cognition.2017.05.031>
- Schiefner, A., Roelofsen, F., & Sümer, B. (2026). Iconicity and transparency in Dutch Sign Language. *Behavior Research Methods*, 58(87). <https://doi.org/10.3758/s13428-026-02969-3>
- Watkins, F., Abdllkarim, D., Winter, B., & Thompson, R. L. (2024). Viewing angle matters in British Sign Language processing. *Scientific Reports*, 14(1), 1043. <https://doi.org/10.1038/s41598-024-51330-1>
- Yang, X., Kukreja, D., Pinkus, D., Sagar, A., Fan, T., Park, J., Shin, S., Cao, J., Liu, J., Ugrinovic, N., Feiszli, M., Malik, J., Dollár, P., & Kitani, K. (2025). SAM 3D Body: Robust full-body human mesh recovery. arXiv Preprint. <https://doi.org/10.48550/arXiv.2602.15989>

Thank you!

... and thank you to my collaborators in this project and the signers who contributed to the database.



<https://doi.org/10.21942/uva.c.8233876.v3>