Translation from spoken to sign languages: state of the art and future challenges

Fabrizio Nunnari, Ph.D.
Senior Researcher at the
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
fabrizio.nunnari@dfki.de

14.05.2025

Sign Languages: basics

- Universal? → No. There are many sign languages
- Text? → No. Only SL is the mother language
- Strong sense of community → “Nothing about us without us”

Sign Languages: communication channels

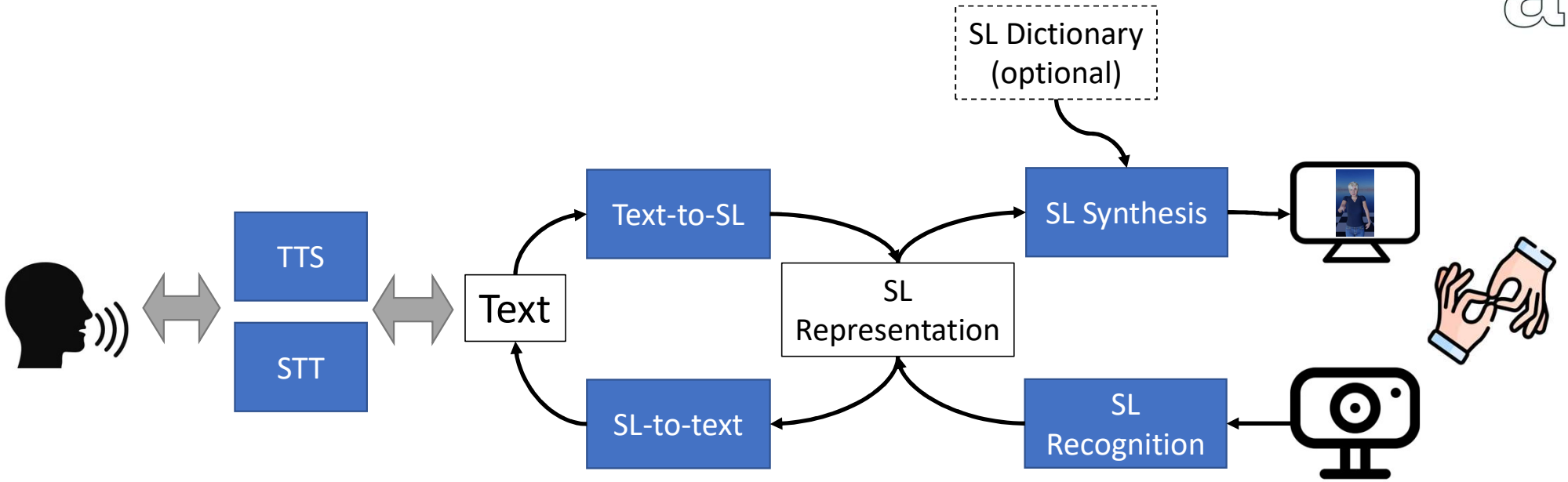
- Hands
  - Hands motion
  - Finger configuration
  - Hands orientation
- Non-manuals
  - Head motion / position
  - Shoulders
  - Torso motion / orientation
  - Facial expressions
  - Mouthings / mouth gestures

DE: Die Bremer Stadtmusikanten. Es war einmal: So fangen Märchen an. Ein Märchen ist eine sehr alte Geschichte.
EN: The Town Musicians of Bremen. Once upon a time: that's how fairy tales begin. A fairy tale is a very old story.

Sign Languages: some features

- Signing space
- Relocation
- Iconicity
- Incorporation
- Role-taking

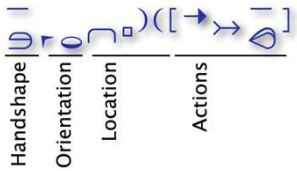
Automated translation



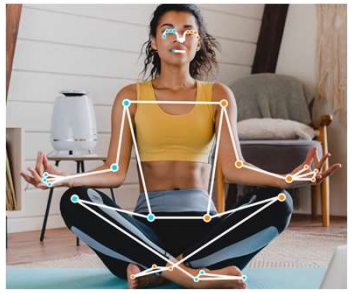
SL intermediate representations

TITEL BREMEN STADT MUSIK $INDEX-dem
MÄRCHEN VOR-LANGER-ZEIT SEHR ALT MÄRCHEN

- GLOSSES: list of words
- HamNoSys: symbols for hand/face configurations [Hanke, 2004]
- SignWriting: symbols for hand/face configurations [Sutton, 1996]
- Animation data:
  - Body skeletal motion / Face muscle activation
- MMS (self-promotion)



cry



mainglos	framestart	frameend	duration	transition	domgloss	ndomgloss	torsorelocx	torsorelocy	torsorelocz	torsorelocax	torsorelocay	torsorelocaz	domhandrelocx	domhandrelocy	domhandrelocz	domhandre
NICHT	0	0	0.5	0.5			0	0	0	0	0	0	0	0	0	
NICHT	0	0	0.5	0.5			-10	0	0	0	0	0.55	0	15	0	
NICHT	0	0	0.5	0.5			5	0	0	0	0	-0.2	0	15	0	

Datasets (1/2)

- PHOENIX (9h, ~9k sentences)
  - Limited domain weather forecast
  - Formal language
  - Single subject signs to camera

[Camgoz et al, 2018]

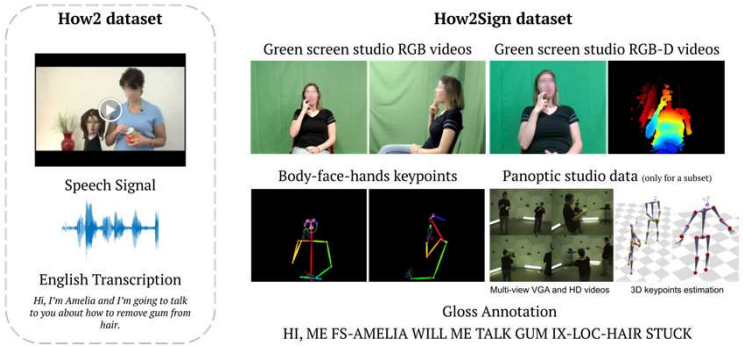
- DGS Korpus (public: 50h, ~25k sentences; private: 560h)
  - Many domains
  - Sitting comfortable position
  - Slang and dialects
  - Interaction between deaf people

[Hanke et al, 2020]

- The Sign Language Dataset Compendium
  - 43 corpora, 86 dictionaries, 28 tasks, 82 signed languages
  - https://www.sign-lang.uni-hamburg.de/lr/compendium/

Datasets (2/2)

- HowToSign
  - American Sign Language
  - 80 hours
- MS-ASL
  - 25000 annotated videos
  - 200 signers
  - 1000 signs
  - Privacy issues!!!



[Duarte et al., 2021]

[Vaezi and Köller, 2019]

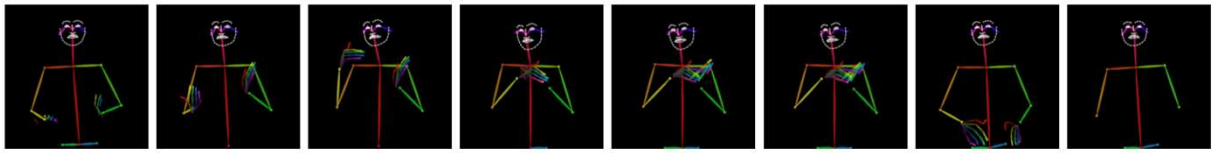
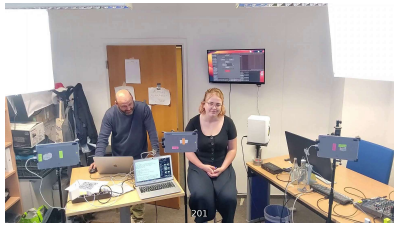


Figure 3: Extracted 137 body key-points for a video sample from MS-ASL by [9, 57].

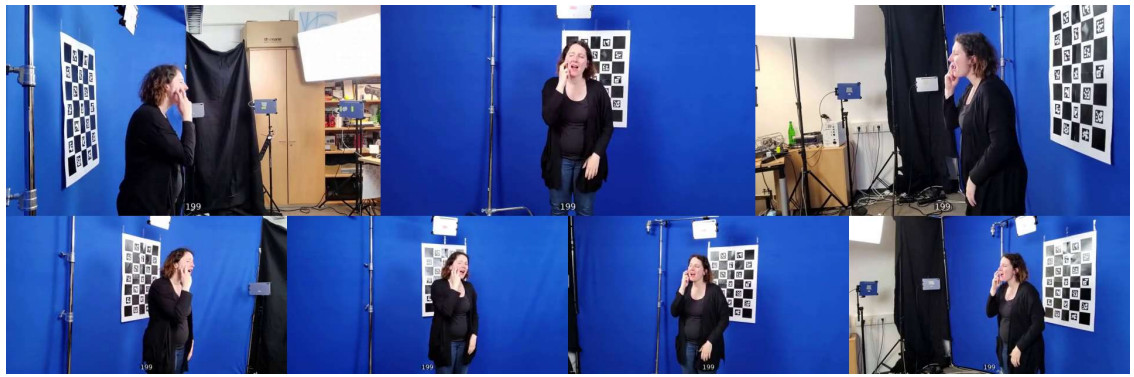
- Token comparison (size issue!!!):
  - DGS private (560h): 657.000 tokens
  - Czech to Ukranian: 115 million

DGS-Fabeln: fairy tales in expressive German sign language



DGS-Fabeln-1 and -2

- Corpus of German fairy tales
- Text // video parallel corpus
- 15 fairy tales
- About 4 hours of recording
- 1400 sentences
- Segmented in ~7s long clips
- High-quality data
- 7+1 points of view
- Suitable for:
  - Translation tasks test suite
  - 3D motion reconstruction
  - Affect analysis
- https://doi.org/10.5281/zenodo.10822096

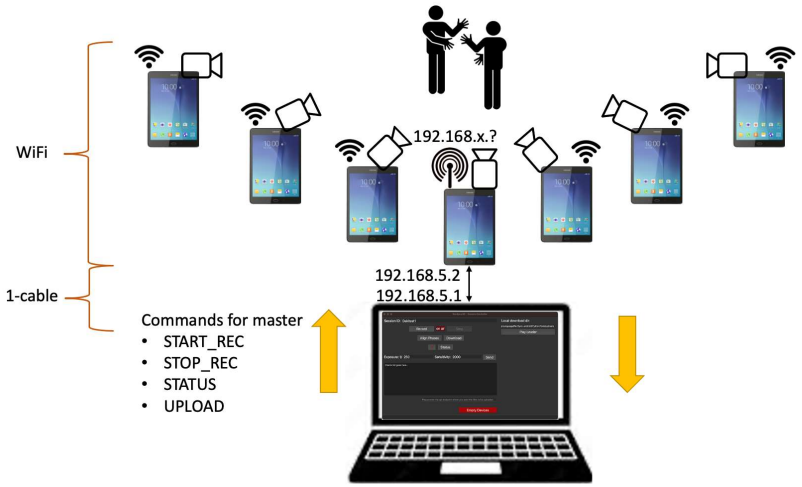


[Nunnari et al., 2024]

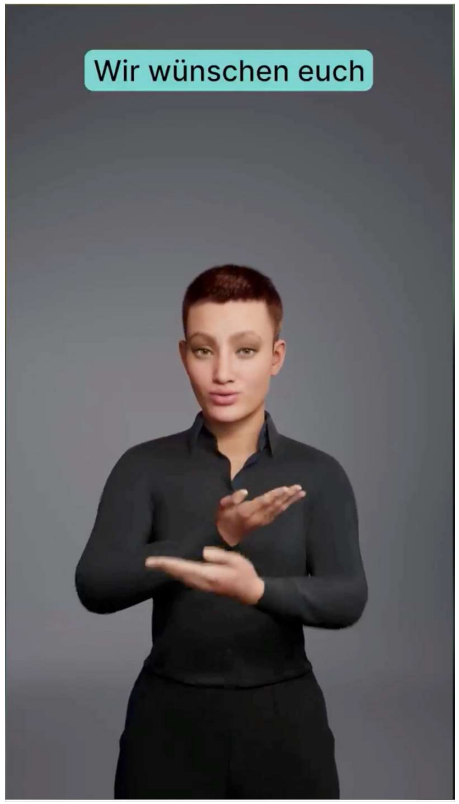
Multi-angle synchronized video recording

RecSyncNG

- Using multiple affordable Android devices for synchronized video capture
- Cable-less sub-millisecond synchronization
- Home-made management GUI
- Open Source: https://github.com/DFKI-SignLanguage/RecSyncNG



SL Avatars: commercial (1/2)



Charamel / Alangu (DE)
https://charamel.com

SignTime (AT) / SiMAX
https://signtime.media

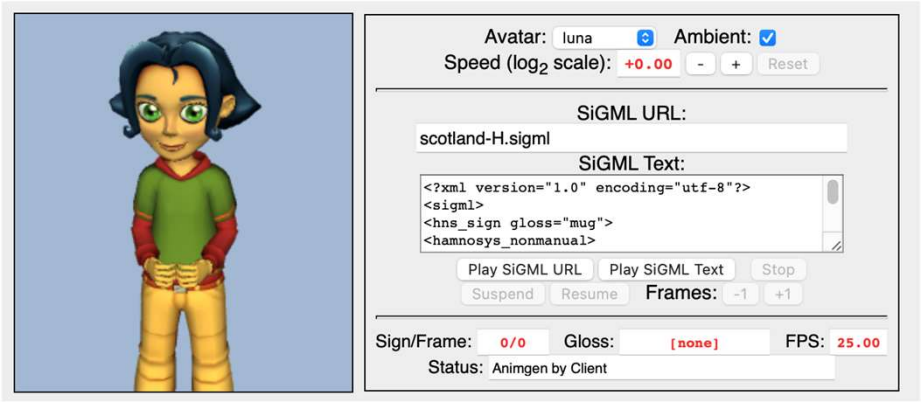
SL Avatars: commercial (2/2)

Migam (US / PL)
https://migam.ai

Signapse (UK)
https://www.signapse.ai

SL Avatars: research (1/2)

- JASigning
  - Uni East Anglia, UK
  - [Kennaway et al., 2007]
- Paula
  - DePaul Univ, Chicago, USA
  - [McDonald et al. 2016]



SL Avatars: research (2/2)

- ATLAS
  - http://www.crit.rai.it/CritPortal/progetti/?p=2566
  - [Vendrame et al., 2010]
- Evolution and Trends in Sign Language Avatar Systems: Unveiling a 40-Year Journey via Systematic Review
  - 47 systems reviewed
  - [Aziz and Othman, 2023]

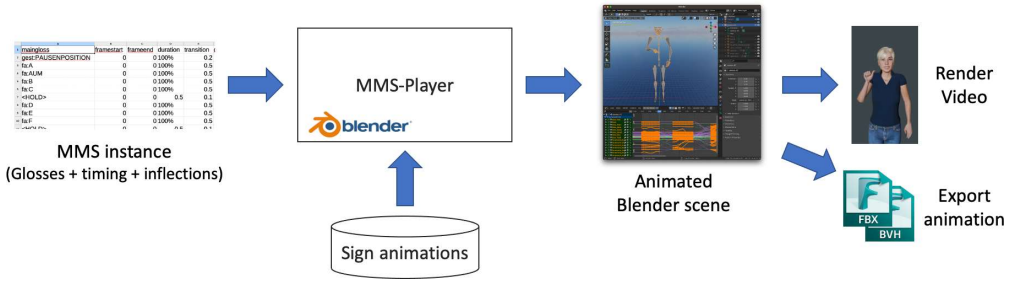
Table 5. Cont.

Ref.	Title	Year	Publication Venue	Cities	Avatar
[74]	Teaching ASL Signs Using Signing Avatars and Immersive Learning in Virtual Reality	2020	conference	1	
[75]	Sign Language Generation System Based on Indian Sign Language Grammar	2020	journal	1	
[76]	3d avatar approach for continuous sign movement using speech/text	2021	journal	3	
[77]	Automatic generation of a 3D sign language avatar on AR glasses given 2D videos of human signers	2021	conference	1	
[78]	SignPose: Sign Language Animation Through 3D Pose Lifting	2021	conference	1	
[79]	An automatic machine translation system for multi-lingual speech to Indian sign language	2022	journal	1	
[80]	Holographic sign language avatar interpreter: A user interaction study in a mixed reality classroom	2022	journal	1	

MMS Player

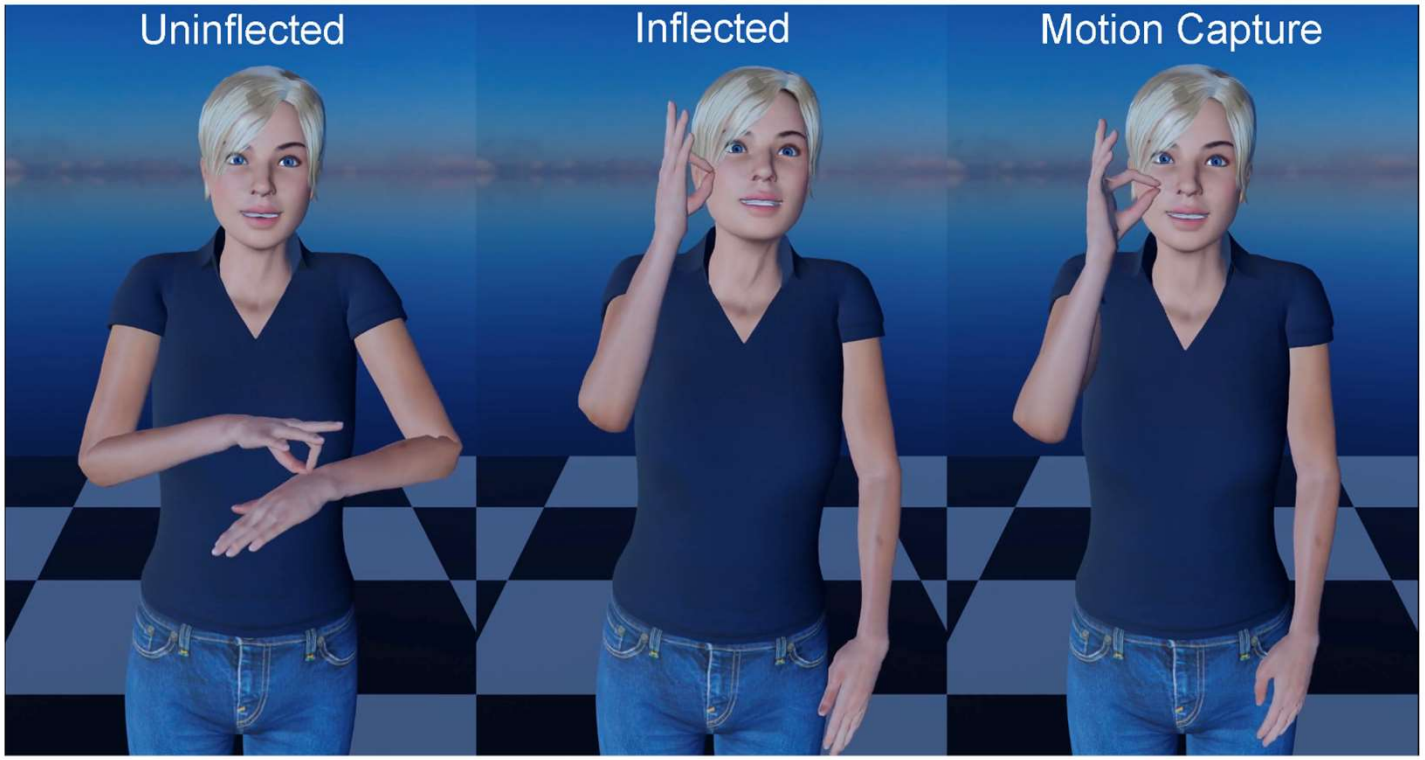
https://github.com/DFKI-SignLanguage/MMS-Player
https://www.youtube.com/watch?v=MxiFsxXukZw

- Procedural animation of Sign Language avatars
- A Blender add-on
  - Reads MMS files
  - Produces SL animations (MP4, FBX, BVH, …, blender scenes)
- MMS: "Multimodal SignStream”
  - GLOSSES + timing + inflections



MMS: The power of inflections

- E.g.: MINUTEN



MMS: timing control

	maingloss	framestart	frameend	duration	transition	domgloss	ndomgloss
2	INDEX	0	0	100%	0.2		
3	INDEX	0	0	30%	0.2		
4	INDEX	0	0	300%	0.2		
5	INDEX	0	0	0.1	0.2		
6	INDEX	0	0	0.2	0.2		
7	INDEX	0	0	1.0	0.2		

MMS: holding a sign

	maingloss	framestart	frameend	duration	transition	domgloss	ndomgloss
2	INDEX	0	0	0.2	0.2		
3	<HOLD>	0	0	0.1	0.1		
4	INDEX	0	0	0.2	0.4		
5	INDEX	0	0	0.2	0.2		
6	<HOLD>	0	0	0.3	0.1		
7	INDEX	0	0	0.2	0.2		
8	INDEX	0	0	0.2	0.2		
9	INDEX	0	0	0.2	0.2		

MMS: finger alphabet example

	maingloss	framestart	frameend	duration	transition	domgloss	ndomgloss
2	gest:PAUSENPOSITION	0	0	100%	0.2		
3	fa:A	0	0	100%	0.5		
4	fa:AUM	0	0	100%	0.5		
5	fa:B	0	0	100%	0.5		
6	fa:C	0	0	100%	0.5		
7	<HOLD>	0	0	0.5	0.1		
8	fa:D	0	0	100%	0.5		
9	fa:E	0	0	100%	0.5		
10	fa:F	0	0	100%	0.5		
11	<HOLD>	0	0	0.5	0.1		
12	fa:G	0	0	100%	0.5		

Augmenting Glosses With Geometrical Inflection Parameters For The Animation Of Sign Language Avatars

F. Nunnari, S. Mishra, and P. Gebhard, “Augmenting Glosses with Geometrical Inflection Parameters for the Animation of Sign Language Avatars,” in 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), 2023, pp. 1–5. doi: 10.1109/ICASSPW59220.2023.10193227.

Problem

- There is not practical written form of SL
- There is no consensus on the “best” formalism
- Researchers rely on GLOSSES, but:
  - Signs can correspond to different words
  - A word can have different signs
  - Some words have no corresponding sign
  - Some signs have no corresponding words
  - GLOSSES do not express timing and spatial variations
    - Some extended-gloss formalism put a patch to it, but it is never enough

Goals for a SL representation format

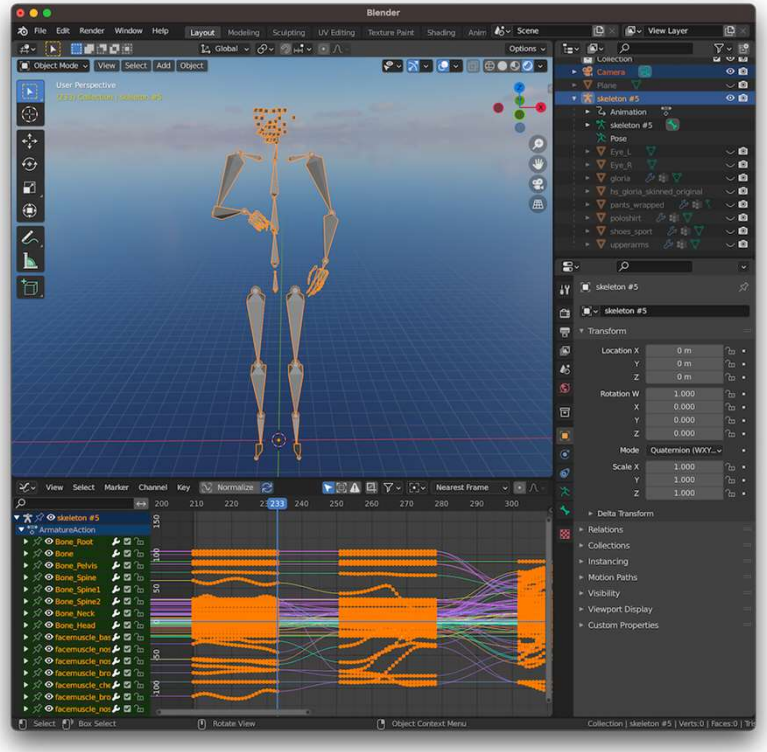
1. A format that does not operate at a linguistic level, but rather at a “geometrical” level
   - Below semantics
   - Applicable to any SL
2. Generalize to the level of casual gesturing
   - Able to describe the manifestation of non-linguistic features like emotional prosody
3. Annotation doesn’t need human observers, but can be algorithmically computed by comparing sentences w.r.t. dictionaries
   - Data creation can be automated
4. A format that can be interpreted by human programmers
   - Easy manual mapping to motion transforming functions
   - Easy to correct through dedicated 3D GUIs
5. Allow for an integration of the translation pipeline with computer-assisted translation (CAT) tools
   - Can be integrated into a complete text-to-SL translation+synthesis pipeline

The MultiModal Signstream (MMS)

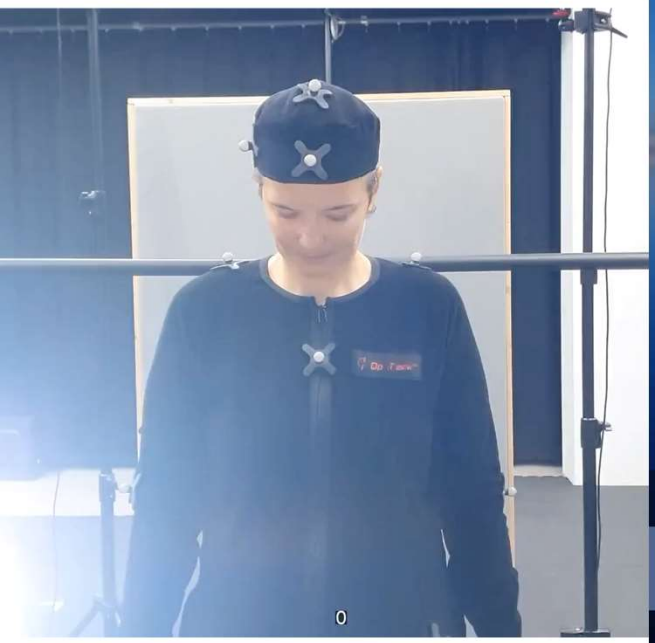
- A table (CSV friendly)
- First column == glosses
- 2X columns: duration and delay
- 2X columns: (non-)dominant hand gloss override (aka Frankenstein)
- 50X+ columns: inflection parameters for
  - Hand trajectories
  - Hands rotation
  - Head rotation
  - Shoulders position
  - Torso position and rotation
  - …

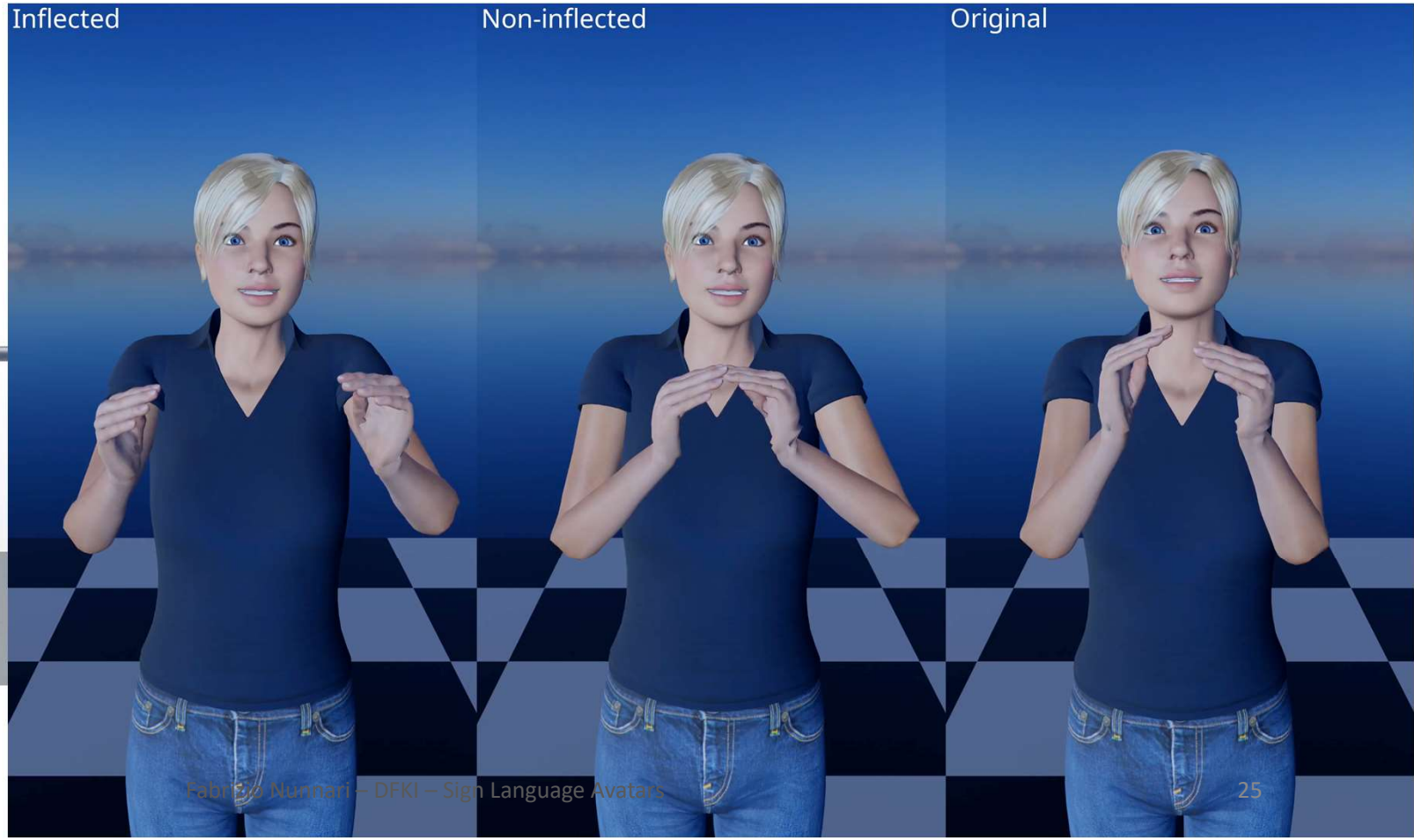
The MMS visualizer

- Blender addon (Open source)
- Compose a sentence:
  - Load a sign
  - Apply inflection
  - Store on timeline
  - Advance cursor
  - Repeat
- Everything through direct coding (no ML here! Yet)



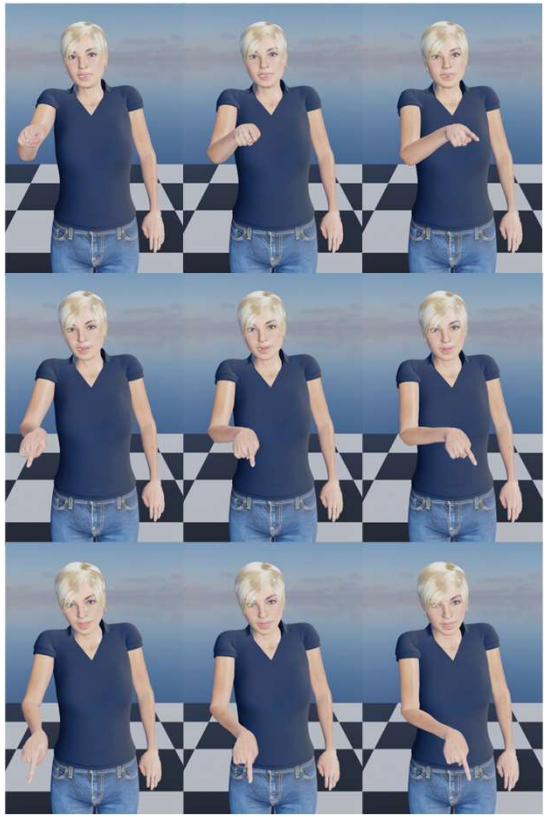
Example: (Satz20) Original + Rendered videos





Examples

INDEX



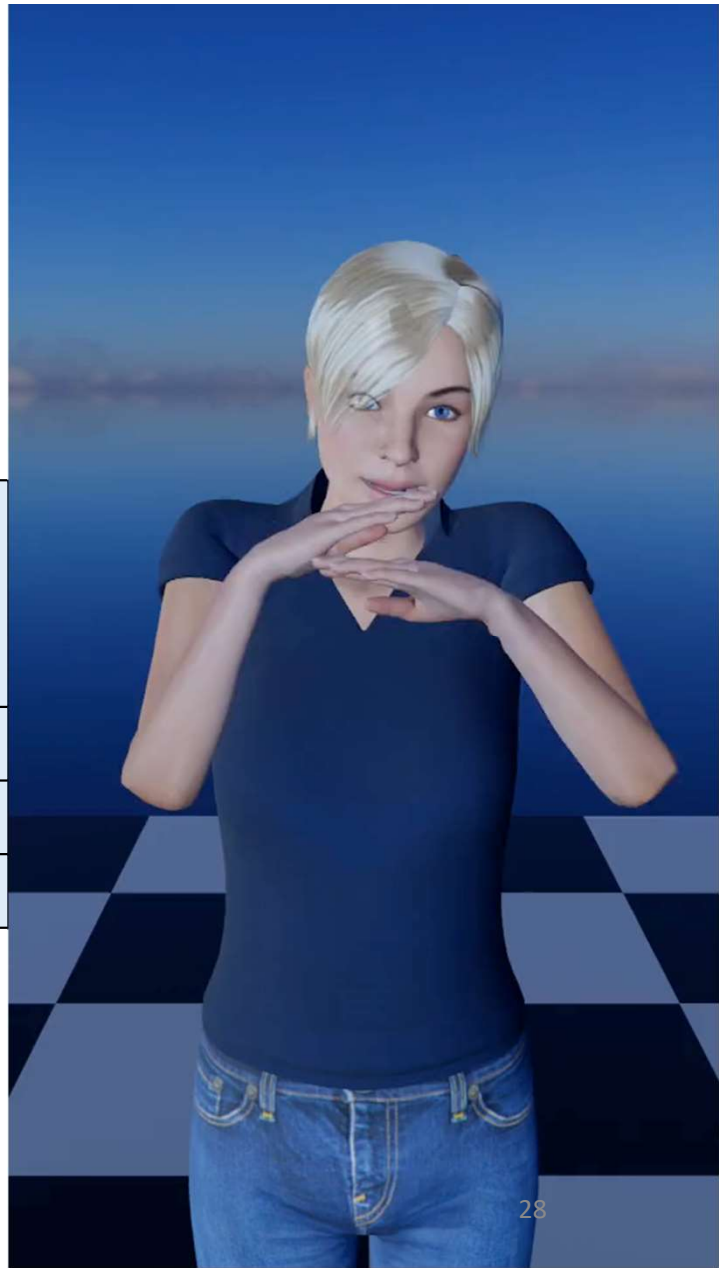
NICHT

Example: INDEX relocated

maingloss	framestart	frameend	duration	transition	domgloss	ndomgloss	domhandrelocx	domhandrelocy	domhandrelocz	domhandrelocax	domhandrelocay	domhandrelocaz	domhandrelocsx	domhandrelocsy	domhandrelocsz	domhandrotx	domhandroty	domhandrotz
INDEX	0	0.2	0.2	0			0	10	-8	0	0	0	1	1	1	0	0.754	0
INDEX	0	0	0.2	0.5			0	10	0	0	0	0	1	1	1	0	0	0
INDEX	0	0	0.2	0.5			0	10	8	0	0	0	1	1	1	0	-0.754	0
INDEX	0	0	0.2	0.5			0	0	-8	0	0	0	1	1	1	0.754	0	0
INDEX	0	0	0.2	0.5			0	0	0	0	0	0	1	1	1	0	0	0
INDEX	0	0	0.2	0.5			0	0	8	0	0	0	1	1	1	-0.754	0	0
INDEX	0	0	0.2	0.5			0	-10	-8	0	0	0	1	1	1	0	0	0.754
INDEX	0	0	0.2	0.5			0	-10	0	0	0	0	1	1	1	0	0	0
INDEX	0	0	0.2	0.5			0	-10	8	0	0	0	1	1	1	0	0	-0.754

Example: NICHT small/mid/large

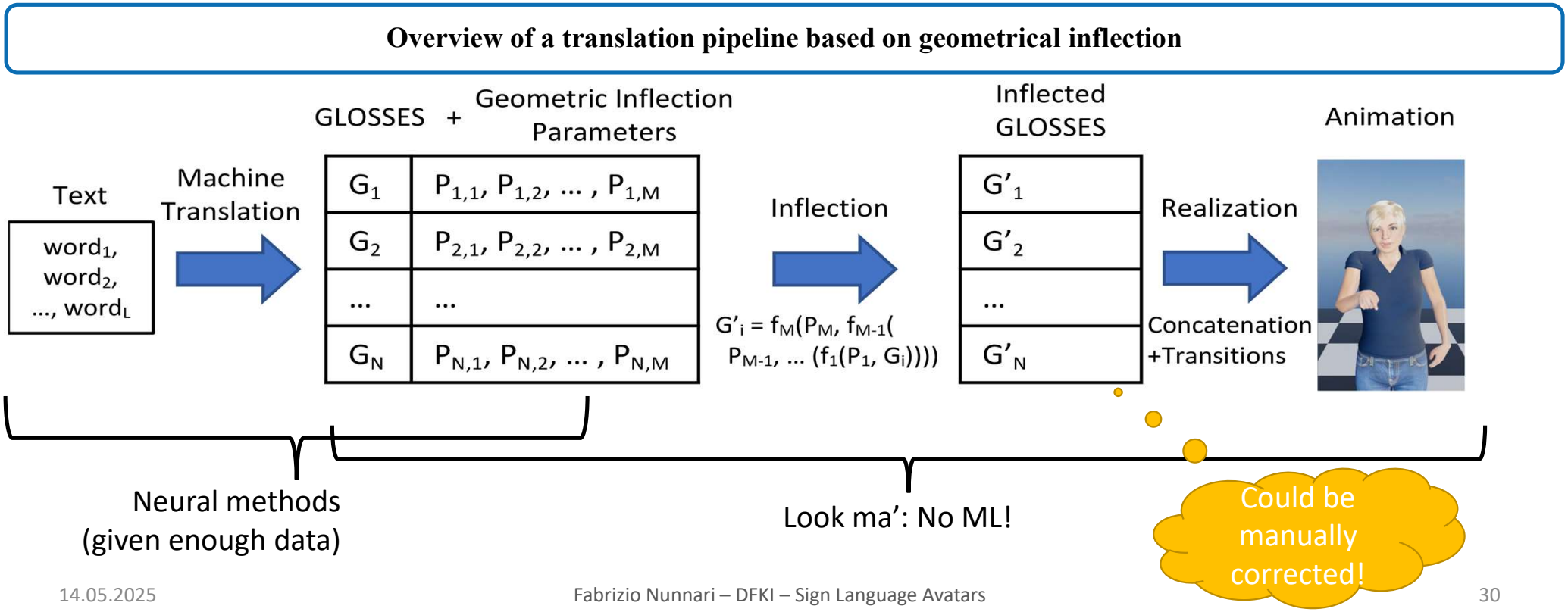
maingloss	frame start	frame end	duration	transition	domgloss	ndomgloss	domhandrelocx	domhandrelocy	domhandrelocz	domhandrelocax	domhandrelocay	domhandrelocaz	domhandrelocsx	domhandrelocsy	domhandrelocsz	ndomhandrelocx	ndomhandrelocy	ndomhandrelocz	ndomhandrelocax	ndomhandrelocay	ndomhandrelocaz	ndomhandrelocsx	ndomhandrelocsy	ndomhandrelocsz
NICHT	0	0	0.7	0.7			0	15	0	0	0	0	0.4	0.4	0.4	0	15	0	0	0	0	0.4	0.4	0.4
NICHT	0	0	0.7	0.7			0	10	0	0	0	0	1	1	1	0	10	0	0	0	0	1	1	1
NICHT	0	0	0.7	0.7			0	5	0	0	0	0	1.8	1.8	1.8	0	5	0	0	0	0	1.8	1.8	1.8



Example: torso inflection (role-shift)

maingloss	framestart	frameend	duration	transition	domgloss	ndomgloss	torsorelocx	torsorelocy	torsorelocz	torsorelocax	torsorelocay	torsorelocaz
INDEX	0	0	0.2	0.5			0	0	0	0	0	0
INDEX	0	0	0.2	0.5			3	0	5	0	-0.78	0
INDEX	0	0	0.2	0.5			-8	0	8	0	0.78	0

Full realization pipeline



Bibliography

- Aziz, M. and Othman, A. (2023) ‘Evolution and Trends in Sign Language Avatar Systems: Unveiling a 40-Year Journey via Systematic Review’, Multimodal Technologies and Interaction, 7(10), p. 97. Available at: https://doi.org/10.3390/mti7100097.
- Camgoz, N.C. et al. (2018) ‘Neural Sign Language Translation’, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- Duarte, A. et al. (2021) ‘How2Sign: A Large-scale Multimodal Dataset for Continuous American Sign Language’, in Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2735–2744.
- Hanke, T. (2004) ‘HamNoSys-representing sign language data in language resources and language processing contexts’, in LREC, pp. 1--6.
- Hanke, T. et al. (2020) ‘Extending the public DGS corpus in size and depth’, in Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives, pp. 75–82.
- Kennaway, J.R., Glauert, J.R.W. and Zwitserlood, I. (2007) ‘Providing signed content on the Internet by synthesized animation’, ACM Transactions on Computer-Human Interaction, 14(3), p. 15. Available at: https://doi.org/10.1145/1279700.1279705.
- McDonald, J. et al. (2016) ‘An automated technique for real-time production of lifelike animations of American Sign Language’, Universal Access in the Information Society, 15(4), pp. 551–566. Available at: https://doi.org/10.1007/s10209-015-0407-2.
- Nunnari, F., Mishra, S. and Gebhard, P. (2023) ‘Augmenting Glosses with Geometrical Inflection Parameters for the Animation of Sign Language Avatars’, in 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), pp. 1–5. Available at: https://doi.org/10.1109/ICASSPW59220.2023.10193227.
- Nunnari, F. et al. (2024) ‘DGS-Fabeln-1: A Multi-Angle Parallel Corpus of Fairy Tales between German Sign Language and German Text’, in Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). Torino, Italy: ELRA and ICCL, pp. 4847–4857. Available at: https://aclanthology.org/2024.lrec-main.434.
- Sutton, V. (1995) ‘Lessons in signwriting–textbook and workbook’, The Center for Sutton Movement Writing, Inc., La Jolla, CA [Preprint].
- Vaezi Joze, H. and Koller, O. (2019) ‘MS-ASL: A Large-Scale Data Set and Benchmark for Understanding American Sign Language’, in The British Machine Vision Conference (BMVC). Available at: https://www.microsoft.com/en-us/research/publication/ms-asl-a-large-scale-data-set-and-benchmark-for-understanding-american-sign-language/.
- Vendrame, M. and Tiotto, G. (2010) ‘ATLAS Project: Forecast in Italian Sign Language and Annotation of Corpora’, in 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. Valletta, Malta, pp. 239--242.

Thank you

Questions?

Contact: <fabrizio.nunnari@dfki.de>