



Maria Kopf, Reiner Konrad, Gabriele Langer, Marc Schulder, Lutz König
Institute of German Sign Language and Communication of the Deaf, University of Hamburg, Germany

Motivation

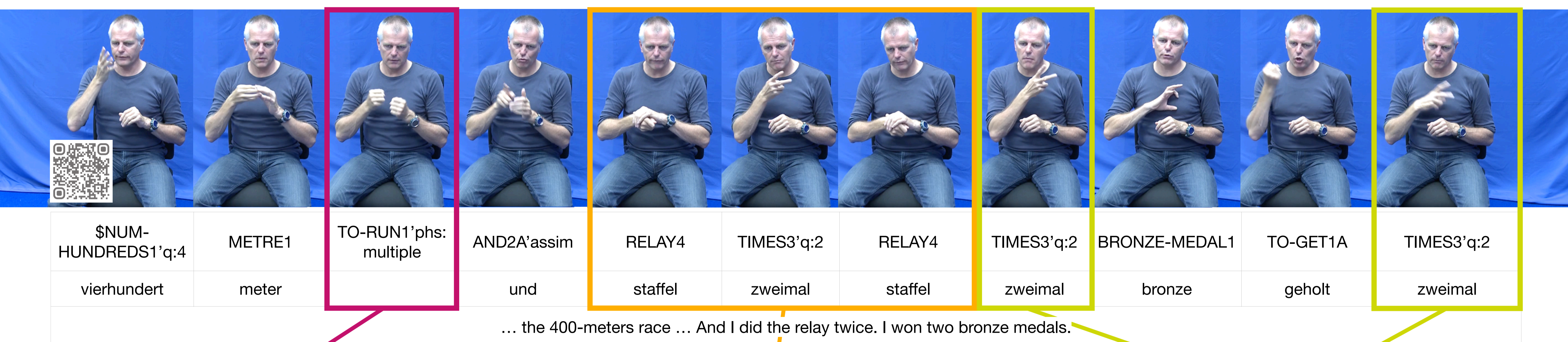
In our research with the DGS Corpus we encounter passages that seem semantically incorrect. Closer inspection shows these to be articulation errors, which are typical in **spontaneous signing**. Since spontaneous signing shares similar challenges with spontaneous speech – studied under the terms **disfluencies**, **turn taking** or **repairs** – we ask: What other phenomena of spontaneous signing occur in our data, and how do we deal with them in our **annotations**? Conversely, how can our existing annotations help detect additional phenomena of spontaneous signing?

Annotation

- Introduction of **two new tiers**: *Disfluency_types* and *Disfluency_parts*.
- Tags in *Disfluency_types* can cover multiple signs
- Tags in *Disfluency_parts* label tokens within repetitions and repairs
- One sequence can display several disfluency types
- **Multiple selection** of disfluency types possible
- Not certain if applicable to other sign languages
- Offers **ideas** and **perspectives** to be tested for other data

Future work

- Dedicated annotation of the presented phenomena
- Scaling for frequency-based analyses beyond the scope of the DGS-Korpus project.
- Future projects hopefully use the DGS Corpus to explore such matters
- Kinematic data gained from automatic pose recognition models
- Revisit the assumptions and paradigms of current annotation schemata.



\$NUM-HUNDREDS1'q:4	METRE1	TO-RUN1'phs: multiple	AND2A'assim	RELAY4	TIMES3'q:2	RELAY4	TIMES3'q:2	BRONZE-MEDAL1	TO-GET1A	TIMES3'q:2
vierhundert	meter		und	staffel	zweimal	staffel	zweimal	bronze	geholt	zweimal

... the 400-meters race ... And I did the relay twice. I won two bronze medals.

Lengthening

Features:	Executions with longer durations [1]
Search:	Query: deviation in token length – above mean value
Types:	Repetition of sign, hold in first or last place of articulation
Edge cases:	Intentional intensifying use, illustrative function

Repair

Features:	Production is halted and continued with correction or new phrase (i.e. false start) [2]
Search:	Queries: combining break-offs in translation and tokens, missing repetition and reoccurring signs
Types:	Repairs of articulation errors; repairs with deletion, insertion, substitution of incorrect signs or lexical variants
Edge cases:	Political correctness, rich regional variation (strategy to ensure understanding)

Repetition

Features:	Basic meaningful elements appear more than once [3]
Search:	Query: neighbouring tokens
Types:	Adjacent, non-adjacent, framing
Edge cases:	Lexical doubling, rich regional variation (strategy to ensure understanding), illustrative function

Pause

Features:	Interruption of information production [6,7]
Search:	Existing annotation: gestures, e.g. \$GEST-OFF1^ (palm-up gesture) or \$GEST-TO-FOLD-ONES-HANDS (folding hands) Query: Long periods with no token tags (<= 1 second)
Types:	Unfilled (no activity of hands and face) and filled (crossed hands, long transition, prolonging of sign segments, gestures without meaning)
Edge cases:	Constructed Action, facial expression used to comment on signing

Elision

Features:	Element or word not realised [8,9]
Search:	Existing annotation: qualifier for movement [10] Query: deviation in token length – below mean value
Types:	Deletion of movement, deletion of whole sign
Edge cases:	Change of Place of articulation, handshape or orientation; artefacts of annotation; illustrative function

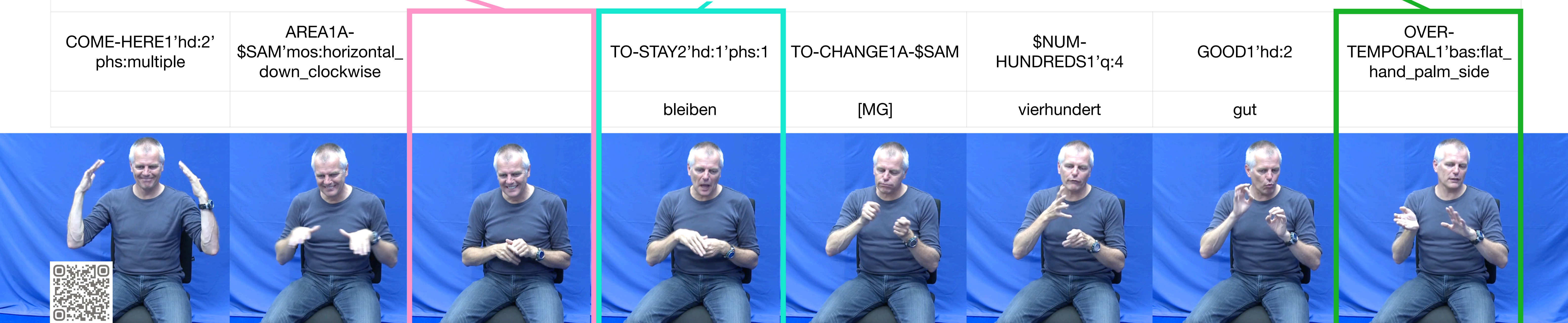
Articulation Error

Features:	Incorrect signing in otherwise correct production [4]
Search:	Existing work: non-tokens [5]
Types:	Anticipations, perseverations, substitutions, blends, fusion, etc. on sub-lexical and lexical level
Edge cases:	Assimilations, signing style, lax signing

Discourse Marker

Features:	Signs that are used on meta-discursive level to structure discourse [11]
Search:	Existing annotation: frequency
Types:	Palm-up, DONE, THEN1A, EQUAL1A, buoys, MEANING1, and others

The audience came and the 400-meters race went well.



COME-HERE1'hd:2' phs:multiple	AREA1A-\$SAM'mos:horizontal_down_clockwise	TO-STAY2'hd:1'phs:1	TO-CHANGE1A-\$SAM	\$NUM-HUNDREDS1'q:4	GOOD1'hd:2	OVER-TEMPORAL1'bas:flat_hand_palm_side
		bleiben	[MG]	vierhundert	gut	

References:
 [1] E. Corblin, et al. 2004. Towards an inclusive system for the annotation of (dis)fluency in typical and atypical speech. In *Clinical Linguistics & Phonetics*, 38(4):381-398.
 [2] E. Shriberg. 2001. To 'error' is human: ecology and acoustics of speech disfluencies. In *Journal of the International Phonetic Association*, 31(1):153-169.
 [3] I. Notarigo & L. Meurant. 2019. Conversations spontanées en langue des signes de Belgique francophone (LSFB) : fonctions et usages de la répétition. In *Lidil*, 60.
 [4] J. Noll, et al. 2003. Sign language production: slips of the hand and their repairs in German Sign Language (DGS). In *ISLS 2003*, pp. 307-303.
 [5] G. Langer, et al. 2016. "Non-tokens": When tokens should not count as evidence of sign use. In *sign-lang/LREC2016*, pp. 137-142.
 [6] S. Gabarró-López & L. Meurant. 2016. Slicing your SL data into basic discourse units (BDUs). Adapting the BDU model (syntax + prosody) to signed discourse. In *sign-lang/LREC2016*, pp. 61-68.
 [7] A. Oviolo, et al. 2018. Zur Bedeutung von Pausen in der Deutschen Gebärdensprache – eine linguistische Annäherung. In *DAS ZEICHEN*, 108:118-125.
 [8] P. Blache, et al. 2017. The corpus of interactional data: A large multimodal annotated resource. In *Handbook of Linguistic Annotation*, pp. 1323-1358.
 [9] L. Corblin. 2018. Discourse Markers and (Dis)fluency Forms and Functions across Languages and Registers. In *Pragmatics & Beyond New Series*, 298.
 [10] C. Looe & R. Konrad. 2022. Detailed annotation and qualifiers. Project Note AP03-2021-01. DGS-Korpus project.
 [11] L. Corblin. 2017. Discourse markers and (dis)fluency in English and French: Variation and combination in the DisFluEn corpus. In *International Journal of Corpus Linguistics*, 22(2):242-269.