HNS2CF
A MAPPING TOOL FROM HAMNOSYS TO SL CATFORM

Lisa Loy
Universität Hamburg
lisa.loy@uni-hamburg.de

Hope E. Morgan
Universität Hamburg
hope.morgan@uni-hamburg.de

LREC 2026 Palma

OBJECTIVE

Many different ways of representing sign forms – e.g., Stokoe Notation [1], SignWriting [2] – have been designed over the past 60+ years, each developed with certain goals and uses in mind. To promote annotation flexibility while avoiding data isolation, we argue there should be a robust effort to translate between systems. Here, we take a first step in this direction by mapping from the Hamburg Notation System (HamNoSys; [3]) to the new SL CatForm coding schema [4].

Our contributions:

- The HNS2CF mapping tool, the formal mapping rules, and implementation
- The Python code used to build the tool
- An empirical evaluation on German Sign Language (DGS) data

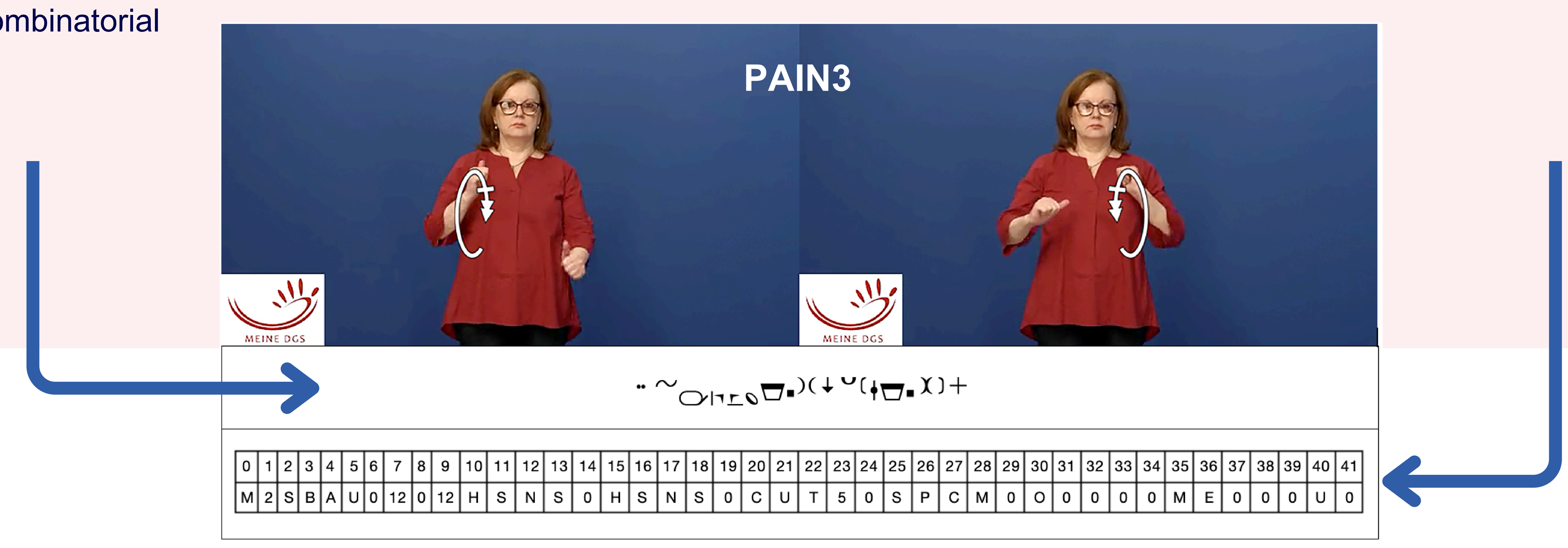
THE TWO NOTATION SYSTEMS

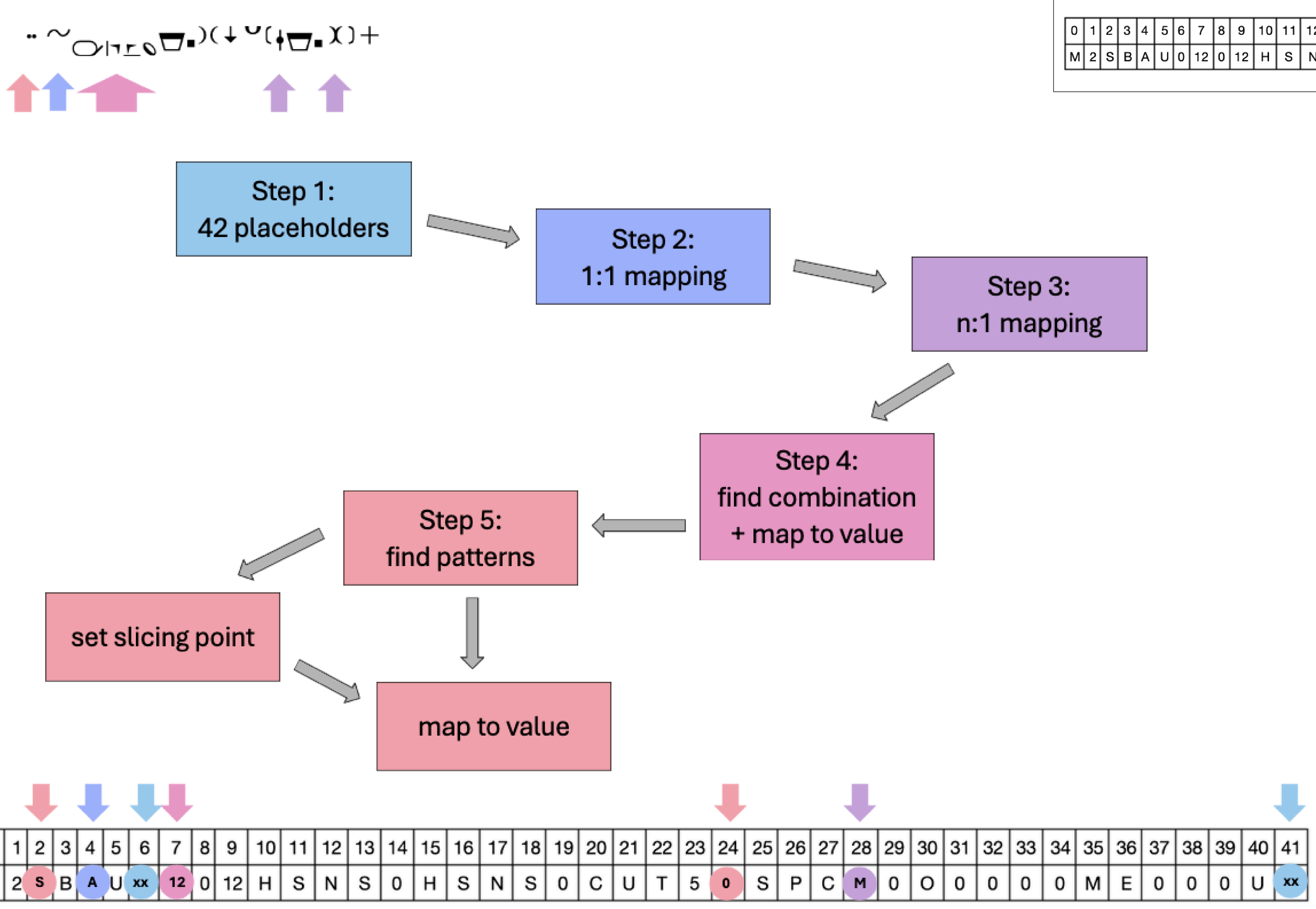
Hamburg Notation System (HamNoSys)

- Originally developed in the 1980s [5], but has undergone several improvements and extensions; we use HamNoSys version 3 [3]
- Encodes signs as linear strings of iconic symbols, representing the manual part of a sign and (optionally) non-manual features
- Encompasses roughly 200–210 characters (see example below [6])
- Follows a specific syntax and is inherently combinatorial

SL Cat Form Coding Schema

- A more recent phonological coding schema, designed to represent minimal units of categorical form in sign languages [4]
- Builds on the descriptive analysis of Kenyan Sign Language (KSL)
- Contains 42 variables, each consisting of a single or double alphanumeric character/s (see example below)
- Variables are fixed and independent => the system does not employ sequential ordering





MAPPING PROCESS

One to one mapping (1:1)

- Most straightforward mapping: direct matches from one HamNoSys character to one value in SL CatForm
- 18/40 variables that were mapped

Many to one mapping (n:1)

- Multiple characters in the HamNoSys string map to a single value within a variable in the SL CatForm string
- The most common mapping relation: 25/40 variables that were mapped

Combination

- Matching value for a variable is extracted from a fixed combination of multiple HamNoSys characters
- 11/40 variables that were mapped

Pattern

- Certain combinations of any character from a set of HamNoSys characters
- Once found, this triggers: a) set value for variable, **or** b) set slicing point, split the HamNoSys, and look only in specific part for matches
- 17/40 variables that were mapped

EVALUATION

Results

- Of 4,116 compared values (98 types * 42 variables) 76.7% (3,155 values; green) were matches, 16.9% (699 values; red) were mismatches, 0.6% (24 values; orange) the mapping tool could not find a suitable match, and 1,1% (42 values; yellow) were empty manual annotations
- Least accurate performance for **orientation** (52.1% mismatches); most accurate performance for **articulator** (only 3.9% mismatches)

Discussion

- First version of HNS2CF already performs well in the evaluation study, with 76.7% matches in 98 random DGS signs
- Many mismatches are not caused by HNS2CF itself, but other sources, e.g., the manual coder not feeling confident or mismatches for finger orientation that were later shown to be mapped correctly from HamNoSys (annotator disagreement)
- Potential improvements: 1) adding language-specific values to SL CatForm; 2) refining mapping logic for path axis 2; and 3) building reversed system from SL Cat-Form to HamNoSys to serve as validation tool and offer insights into the kind and amount of information lost during mapping

	matches	mismatches	no match found	no manual coding	not possible
articulator	561 (81.8%)	27 (3.9%)	0 (0%)	0 (0%)	98 (14.3%)
handshape	987 (77.5%)	265 (20.8%)	22 (1.7%)	0 (0%)	0 (0%)
orientation	102 (52.1%)	91 (46.4%)	0 (0%)	3 (1.5%)	0 (0%)
location	306 (78.1%)	80 (20.4%)	0 (0%)	6 (1.5%)	0 (0%)
core movement	684 (77.6%)	177 (20.1%)	2 (0.2%)	19 (2.2%)	0 (0%)
manner of movement	515 (75.1%)	59 (8.6%)	0 (0%)	14 (2.1%)	98 (14.3%)
total	3155 (76.7%)	699 (16.9%)	24 (0.6%)	42 (1.1%)	196 (4.8%)

For more information, please visit https://semasign.eu

Funded/Co-funded by the European Union (ERC, SemaSign, No.101117395). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor thegranting authority can be held responsible for them.

erc European Research Council Established by the European Commission

SemaSign

Universität Hamburg DER FORSCHUNG | DER LEHRE | DER BILDUNG

References:

[1] William C. Stokoe. 1960. Sign language structure: *An outline of the visual communications systems of the American Deaf*. Silver Springs, MD: Linstock Press. **[2]** Valerie Sutton. 1999. SignWriting: On the occasion of its 25th anniversary November 1999. *Sign Language & Linguistics 2.2*. **[3]** Thomas Hanke. 2004. HamNoSys – representing sign language data in language resources and language processing contexts. In *Proceedings of the LREC2004 Workshop on the Representation and Processing of Sign Languages: From SignWriting to Image Processing. Information techniques and their implications for teaching, documentation and communication*, pages 1–6, Lisbon, Portugal. European Language Resources Association (ELRA). **[4]** Hope E. Morgan. 2026. SL CatForm coding schema v.1: A field-based phonological coding for sign languages (Version 1.0). Technical report, Universität Hamburg. **[5]** Siegmund Prillwitz, Regina Leven, Heiko Zienert, Thomas Hanke, Jan Henning, et al. 1987. *HamNoSys. Hamburg Notation System for Sign Language. An Introduction*. Zentrum für deutsche Gebärdensprache, Hamburg, Germany. **[6]** Reiner Konrad, Thomas Hanke, Gabriele Langer, Dolly Blanck, Julian Bleicken, Ilona Hofmann, Olga Jeziorski, Lutz König, Susanne König, Rie Nishio, Anja Regen, Uta Salden, Sven Wagner, Satu Worseck, Oliver Böse, Elena Jahn, and Marc Schulder. 2020. MEINE DGS – annotiert. Öffentliches Korpus der Deutschen Gebärdensprache, 3. Release / MY DGS – annotated. Public Corpus of German Sign Language, 3rd release.

Poster presented at the *12th Workshop on the Representation and Processing of Sign Languages: Language in Motion* at LREC 2026, Palma de Mallorca, Spain. 16 May 2026.