



From Form to Function.

A database approach to handle lexicon building and spotting token forms in sign languages

Reiner Konrad, Thomas Hanke, Susanne König, Gabriele Langer, Silke Matthes, Rie Nishio, Anja Regen University of Hamburg, Institute for German Sign Language and Communication of the Deaf





I. Corpus building

Task: lemmatisation

Issue: no general dictionary of DGS available

Answer: lexical database building

Theoretical challenge: combining top-down and bottom-up approach **Practical challenges**: time, resources, consistency

Solutions: multi-user environment, database approach, two-step procedure: token-type matching and lemma revision, separating token and type information

Details: IDs (types as database entities) instead of ID-glosses, relations (hierarchical and others), built-in restrictions, triggers, customised views on the data (listing, sorting, searching), token-type matching by drag & drop (of existing type into token-tag), access to up-to-date data for all users (impedes offline solution)



Token-type matching



ilex://types.id=6889; sub-type's name = JA1A (yes)

Lemma revision

Ĵ.	Form	Meaning	Glosses	Tokens	∞ Language Stills					
×.	192 Entries									
Hand	Film	Timecode	Mundbild	Bedeutung	HamNoSys					
d	eSIGN012	00:02:28:03	ja							
d	eSIGN016	00:06:10:01	ja							
d	eSIGN025	00:02:42:15	ja		-					
d	eSIGN032	00:05:05:10	ja							
d	GLx19702	00:03:46:24			VØØ					
d	GLx19802	00:01:09:13	ja		· 🕞					
d	F1_L0202	00:00:00:09			+ (N)					
d	F1_L0207	80:00:00:00			+ (N)					
d	F1_L0508	00:00:00:14			+ (N)					

2. Taking iconicity of signs into account

Task: Taking the iconicity of signs into account in lexicology and lexicography **Issue**: dynamic sign-mouthing combinations ("iconic-combinatorial procedure"); (iconic) signs can cover a far wider range of meanings than words; same form, same meaning \rightarrow same sign will lead to mapping the German lexicon onto the DGS lexicon

- **Answer**: form, meaning, and iconic value: same form, same iconic value (+ same image producing technique) \rightarrow same sign; description and analysis of the underlying image of signs
- **Theoretical challenges**: discovering iconic relations between signs (iconic network); lemma selection based on iconic principles
- **Practical challenges**: determining the degree of conventionalisation of signs
- and sign-mouthing combinations (regular and occasional collocations)
- **Solutions**: Type hierarchy and double glossing; type information: form and iconic value (image description and image producing techniques); sub-type information: lexicalised meanings
- **Details**: Matching tokens either to sub-types (= regular collocations/ conventional sign-mouthing combinations) or types ((= occasional collocations/productive sign-mouthing combinations); context menu (right mouse click) showing all existing sub-types of one type





3. Towards a full-form sign language lexical database



Issues: There is neither a descriptive grammar nor a corpus-based dictionary of DGS; descriptions of form variation and inflection often have little empirical



00			Termine from dgskorpus_h							
	3	Timecodes	Lexem/Ge	HamNo	HamNoSys_B	Mundbi	Komm			
	SL F	12:04:37:01 12:04:37:08	\$NUM-UHRI		≝r0 ^{(+→)+}	drei uhr				
		12:04:37:08 12:04:37:11			÷					
		12:04:37:11 12:04:37:17	DA1	DA1	-					
		12:04:37:17		DA1-\$	SAM					

evidence.

Answer: Spotting token forms as an intermediate step to base lemma selection (esp. variation), POS tagging, and sign modification on empirical findings **Theoretical challenge**: Defining categories that help to validate known formfunction units and to discover new ones

Practical challenges: consistency

Solutions: grouping form features as qualifiers with/without feature values; extending type hierarchy by qualified types and qualified sub-types; **Details**: Matching tokens either to qualified types or qualified sub-types; context menu (right mouse click) showing all existing qualified types, sub-types, and qualified sub-types of one type

Qualifiers and feature values (work in progress) * no vocabulary c.v. = closed vocabularyo.v. = open vocabulary Number of hands Sub-system manual alphabet Location Iocation ('loc:vocabulary not yet implemented) number of hands ('hd:c.v.) m \checkmark fa one-handed (**'1**:o.v.) M í ì blocation_time_horizontal ('loc_t_h:c.v.) 11 closed vocabulary: open vocabulary Ľ, blocation_time_sagittal ('loc_t_s:c.v.) • 2 > location_time_vertical ('loc_t_v:c.v.) SY SY • 2x 2rev A UB, location_text structure_horizontal ('loc_ts_h:c.v.) • 2acyc closed vocabulary: P • left • right \$ALPHA'1:T \$ALPHA'1:T ASL ARBEITEN2'loc_ts_h:left ARBEITEN2'loc_ts_h:right leftover KANN1 KANN1'hd:2 unclear (to work) ▶ fa two-handed ('2:o.v.) (can) > location_text structure_sagittal ('loc_ts_s:c.v.) ▶ fa tracing ('sk:o.v.) **Movement** > location_text structure_vertical ('loc_ts_v:c.v.) If a tracing on hand ('skh:o.v.)



Poster presented at the 5th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, 2012 LREC Conference in Istanbul, May 27, 2012. This publication has been produced in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities.

Illustrations: Olga Jeziorsky, Dolly Blanck