

DGS-Korpus project (2009-2023)

Aims

- Reference corpus of German Sign Language (DGS)
- Free public corpus (≈ 50 hrs)
- Corpus-based dictionary DGS – German (DW-DGS)

Reference corpus

- Data collection: 2010-2012
- 330 (near-)native signers
- 13 regions across Germany
- 560 hrs signing, 92.5 hrs lemmatised, 615,000 tokens

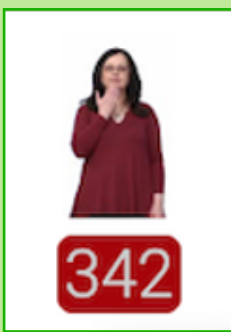
Lexical database (iLex)

- Type hierarchy (double glossing, dependent subtypes)

DW-DGS <https://dw-dgs.de>

Pre-release entries of DW-DGS

- Gloss free: sign representation via moving icons (micons)
- Micons used for cross-references
- Access: DGS index, subject area index, German index (German equivalents)
- Entries include: variants, regional distribution, sense description(s), mouthings, German equivalents, synonyms, antonyms, collocational patterns, examples
- Examples taken directly from the corpus



MEINE DGS <https://meine-dgs.de> <https://ling.meine-dgs.de> <https://annis.meine-dgs.de>

Public DGS Corpus in three formats

- **MY DGS**
 - **Community portal:** subtitled videos
- **MY DGS – annotated**
 - **Research portal:** annotations in German and English
 - Online transcripts
 - Types list: video and HamNoSys of citation form, KWIC view of tokens
 - Download files: videos, annotations, metadata, OpenPose
- **MY DGS – ANNIS**
 - **Corpus query portal** without links to other resources

LSP2 (2007)
Health and Nursing

LSP1 (2010)
Horticulture and Landscaping

Language of Special Purpose (LSP) Dictionaries (1995-2010)

- German technical terms with definitions and DGS equivalents
- Additional index of signs with basic lexicographic description

Benefit of linking

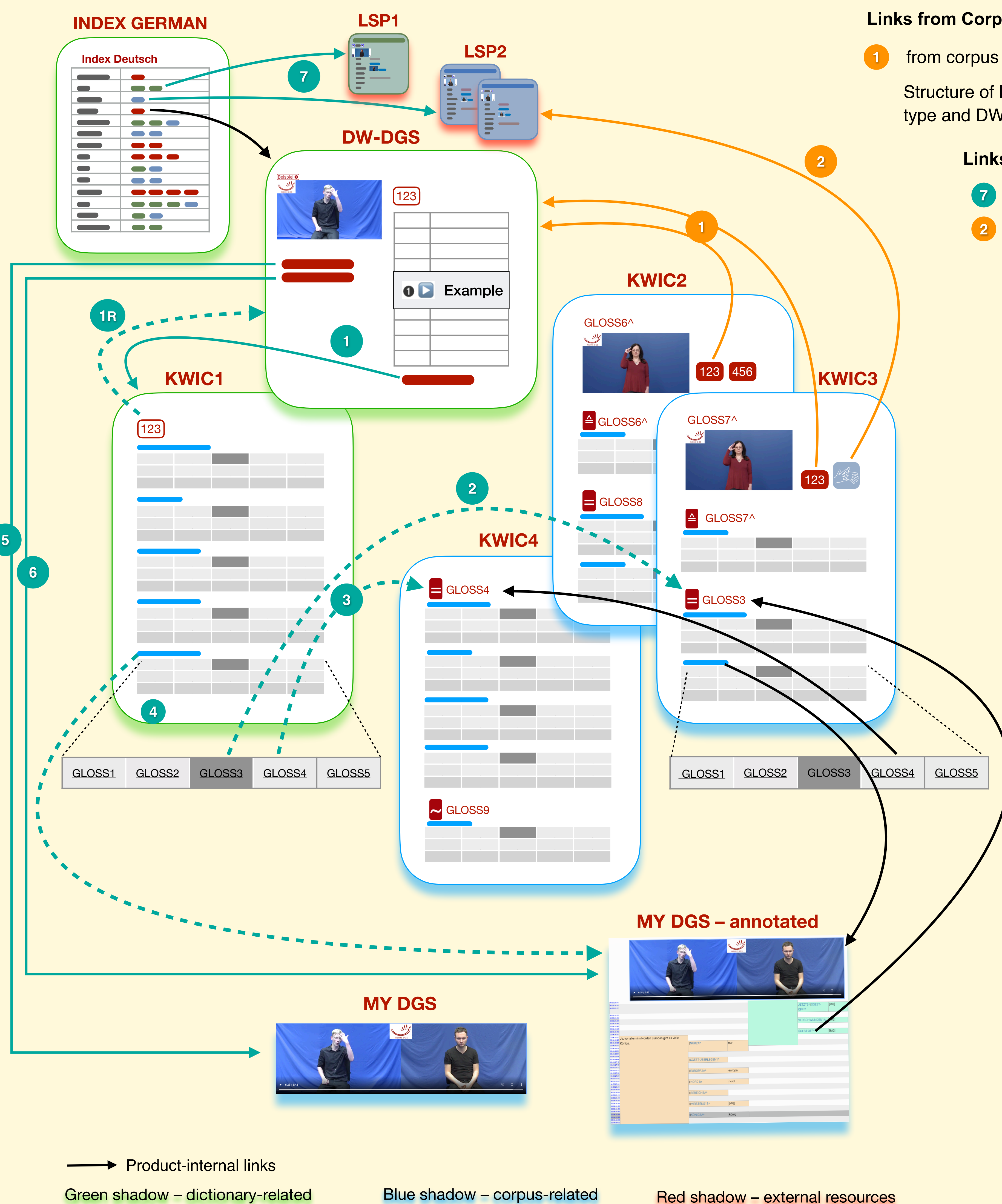
- Supplementary information on signs without DW-DGS entries
- Information not provided in the DW-DGS or in MY DGS – annotated, e.g. iconic motivation of sign

Links from DW-DGS to Corpus

- 1 from DW-DGS entry to KWIC1 concordance of all tokens of all types constituting the entry, and back

The KWIC1 “Keyword in context” concordance is generated from the entry and cannot be accessed from the public corpus directly. It looks similar to a KWIC concordance of corpus types entries (KWIC2-4), but it does not show the type’s video representation or type gloss headers. It shows only the glosses of single tokens.

- 2 from a KWIC1 target token (dark grey shade) to the corresponding corpus KWIC3 or KWIC2
- 3 from any KWIC1 neighbour token to corresponding corpus concordance KWIC4
- 4 from header of a KWIC1 line to source transcript MY DGS – annotated
- 5 from a DW-DGS example sentence to their source text in MY DGS (video) and
- 6 MY DGS – annotated (transcript) respectively



Links from Corpus to DW-DGS

- 1 from corpus types entries KWIC2-4 to DW-DGS entry
- Structure of links depends on matching of corpus type and DW-DGS entry (see Table below).

Links to heterogeneous resources

- 7 from DW-DGS GERMAN INDEX and
- 2 from KWIC2-4 to LSP sign entries

Interlinking Corpus and Dictionary

- usable as stand-alone products
- additional, complementary information and functions
- increases functionality

Challenges

- lemmatisation: scope of types & entries may differ
- interlinking requires explicit form of mapping
- solution: glosses ↔ micons

Linking from Corpus and DW-DGS to LSP Dictionaries

- MY DGS: gateway
- old resources available
- no legal obstacles
- shared database structures in iLex
- mapping and linking from corpus and DW-DGS to LSP dictionaries, not back

Challenge

- Differences in lemmatisation and glossing due to new data, re-evaluation, and change of annotation conventions

Mapping of dictionary entry number and DGS corpus type

Entry in DW-DGS	Relation	Type in MY DGS – annotated	Comment
7	1:1	FISH1^	entry contains all forms of type [icon] [icon]
59	1:2	CROWN1A^ CROWN1B^	entry contains two form variants each belonging to one type [icon] [icon]
128	2:1	TO-PRAY1B^	entry contains base form of type as variant [icon]
615			entry contains modified form of type as variant (lexicalised modification) [icon]

Symbols in Types Entry

- [icon] Conventionalised use of the sign, same form (as type)
- [icon] Conventionalised use of the sign, form is not identical to the sign (modified form of the type)

Further literature:

- Hanke, Thomas / Konrad, Reiner / Jahn, Elena / Schuldner, Marc (2022): “Extending the Public DGS Corpus in Size and Depth”. sign-lang@LREC.
- Isard, Amy / Konrad, Reiner (2022): “MY DGS – ANNIS: ANNIS and the Public DGS Corpus”. sign-lang@LREC.
- Konrad, Reiner, Hanke, Thomas, Langer, Gabriele, König, Susanne, König, Lutz, Nishio, Rie, & Regen, Anja (2022, June 14): Public DGS Corpus: Annotation Conventions / Öffentliches DGS-Korpus: Annotationskonventionen (Version 4.1). <https://doi.org/10.25592/uhhfdm.10251>.
- Langer, Gabriele / Müller, Anke / Wähl, Sabrina / König, Susanne / Hanke, Thomas / Konrad, Reiner (2020): Lemmatisierungsregeln. Vorläufige Wörterbuch-Einträge (Version 4). <https://doi.org/10.25592/uhhfdm.1899>
- Müller, Anke / Langer, Gabriele / Otte, Felicitas / Wähl, Sabrina (in press.): Creating a Dictionary of a Signed Minority Language: A Bilingualized Monolingual Dictionary of German Sign Language. In: Proceedings of the XX EURALEX International Congress, Mannheim, (2022, July 12-16).