The DGS-Korpus approach to including frequent sign combinations in a corpus-based electronic sign language dictionary

Gabriele Langer, Anke Müller, Sabrina Wähl, Thomas Hanke

University of Hamburg, Institute of German Sign Language and Communication of the Deaf

Dictionary
- Preliminary entries, final version to be published in 2023

Corpus-based and Largely Corpus-bound
- Corpus collected 2010-2012, size: 576.400 tokens lemmatised to date
- Type of data: natural and near natural conversations of 330 signers across Germany

Monolingual in General Design
Clear focus on description of German Sign Language (DGS), including information on:
- form variants
- grammatical properties of signs
- regional distribution and differences in use across age groups
- division of senses based on analyses of corpus data
- definition/explanation of meaning; usage information
- example sentences taken directly from the corpus
- synonyms and antonyms (cross-referenced)
- collocational patterns

Some Features of a Bilingual Dictionary
- German translational equivalents included
- Bidirectional access:
  - access by entry number/micon (➜DGS)
  - implemented: search by German equivalent (➜Deutsch)
  - planned: search by sign form

Additional Features
- Metalanguage used for description: German
- Access by topics (➜Sachgruppen)
- Playful access by graphical display (➜Graph)

Design Decision: a Dictionary without Glosses
- Rationale: glosses invite inadequate inferences from gloss word to sign
- Signs should not be represented by words; instead they are represented by micons

Analysis of Frequent Sign Combinations
- Statistics (Mutual Information) helps to find candidates for frequent sign combinations
- In the MI-query phonological and lexical variants are treated as instantiations of the same potential neighbour.
  Reason: fine-grained lemmatisation with regard to form and high lexical variation often result in small numbers of tokens per type
- For lexicographic purposes the focus is not only on specific collocations but also on semantic preferences
- Results provide a rough first impression but need to be evaluated and selected in the analysis stage
- Candidates for frequent sign combinations help to identify different senses of a sign and support the process of Word Sense Discrimination (WSD)

position	left neighbour	base	right neighbour	MI-value	pattern occ	informants dist	candidate occ	base occ	neighbour-glosses
left	DEAF	COMMUNITY1A		5.00	5	3	3496	26	DEAF1A|DEAF1B
left	DEAF	GROUP1A		3.65	17	12	3496	225	DEAF1A|DEAF1B
left	HEARING	GROUP1A		3.38	6	6	1488	225	HEARING1A|HEARING1B
left	MY	GROUP1A		2.67	6	5	2430	225	MY1
left	WARM	TOGETHER1A		6.63	10	8	172	342	WARM1A|WARM1B|WARM2C
left	DEAF	TOGETHER1A		3.28	20	17	3496	342	DEAF1A|DEAF1B
left	AREA	TOGETHER1A		3.26	5	5	886	342	AREA1A|AREA1A–$SAM
left	MORE	TOGETHER1A		2.67	6	6	1603	342	MORE1|MORE3|MORE8B
left	BEAUTIFUL	TOGETHER1A		2.65	5	5	1357	342	BEAUTIFUL1A|BEAUTIFUL1B
left	SCHOOL	TOGETHER1A		2.15	5	2	1918	342	SCHOOL1A|SCHOOL3
right		TOGETHER1A	TO-DRIVE	4.34	5	5	421	342	TO-DRIVE1|TO-DRIVE3
right		TOGETHER1A	TO-WORK	3.44	12	10	1875	342	TO-WORK1|TO-WORK2|TO-WORK2–$SAM
right		TOGETHER1A	BUT	2.36	7	7	2312	342	BUT1
right		TOGETHER1A	$NUM-ONE-TO-TEN	1.16	11	11	8371	342	$NUM-ONE-TO-TEN1A|$NUM-ONE-TO-TEN1B|$NUM-ONE-TO-TEN1D

left neighbours / analysed lexemes / right neighbours — full glosses of neighbours contributing to pattern

Within the query results three different types of sign combinations can be found:
- phrase-like structures
- collocational patterns
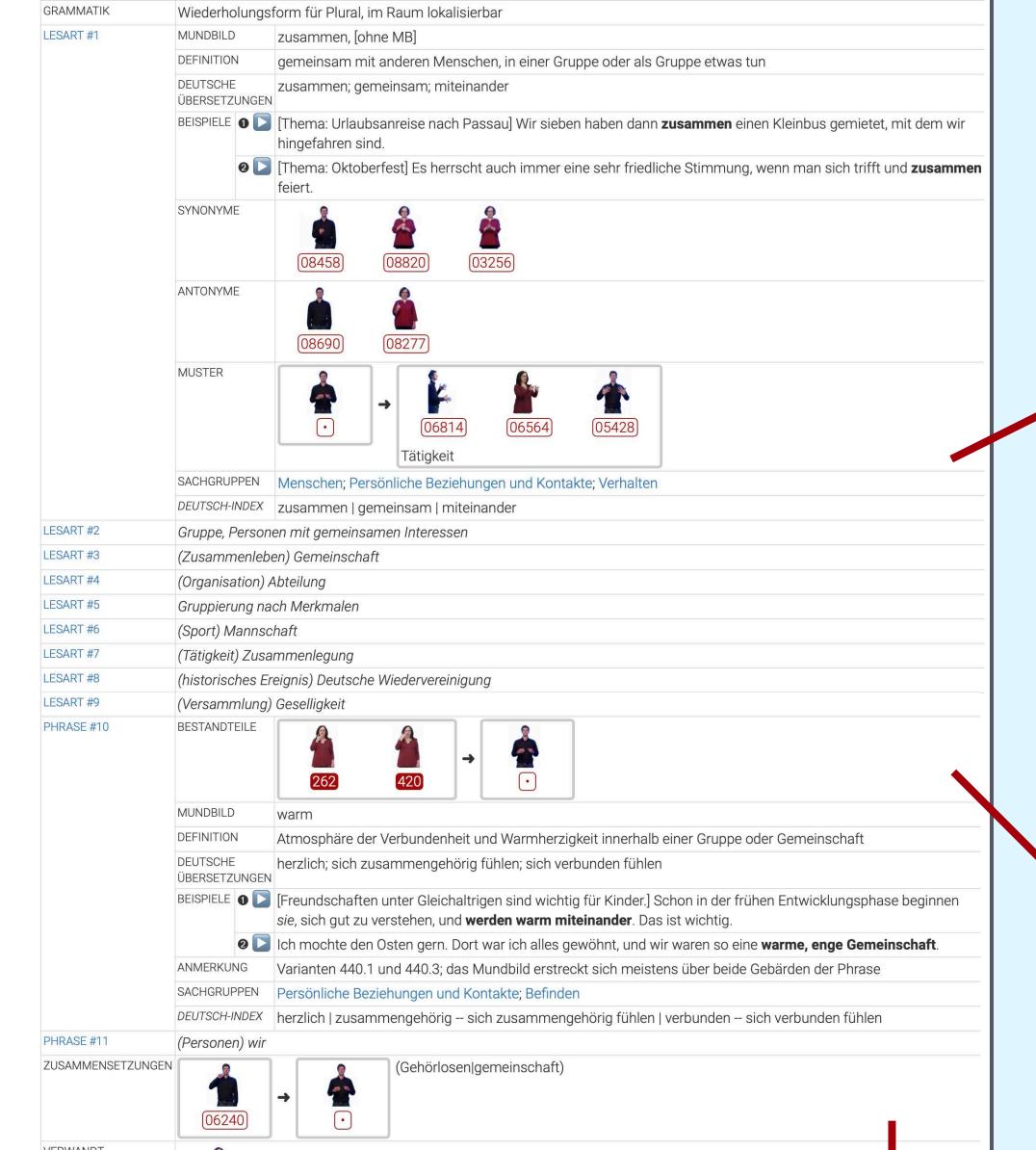- loan compounds and compound-like compositions

Collocational Patterns
- Combinations of frequent neighbours
- Meaning of combination can usually be inferred from individual meanings of both partners
- Theory: One partner, the collocator, is usually less predictable and more restricted as to combinability; it is defined and understood completely only through the base (comp. Svensén 2009, 161-162).
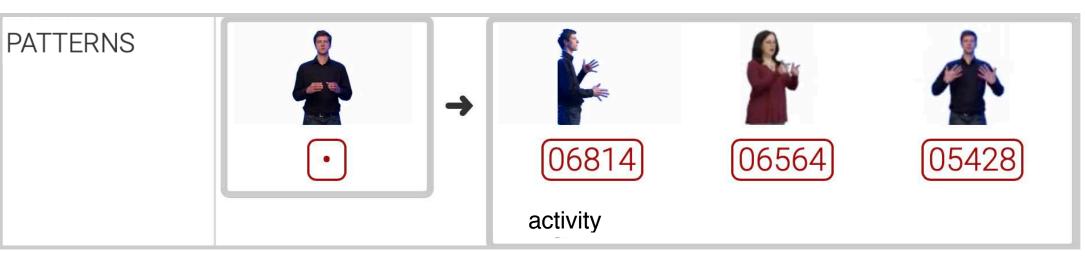- Practice: In our dictionary we also include some selected frequent free combinations in the collocational patterns slot when they are chunks that are useful for clarifying the meaning of the respective sense or its typical usage to the learner.

- Collocational patterns are specific to a sense and therefore addressed as information to the sense.

PATTERNS → 06814 06564 05428

activity

440#1 'together' often co-occurs with signs indicating activities, e.g. 'to sign', 'to drive', 'to play'

Phrases
- DGS-specific combinations of signs that as a whole have a separate meaning, usually not inferable from the meanings of the individual parts of the phrase or surpassing their combined meanings (principle of idiomaticity)
- One mouthed word or mouth gesture often stretches across both partners

- The combined meaning of the signs is not transparent. Thus it is represented as a separate sense in the entry.

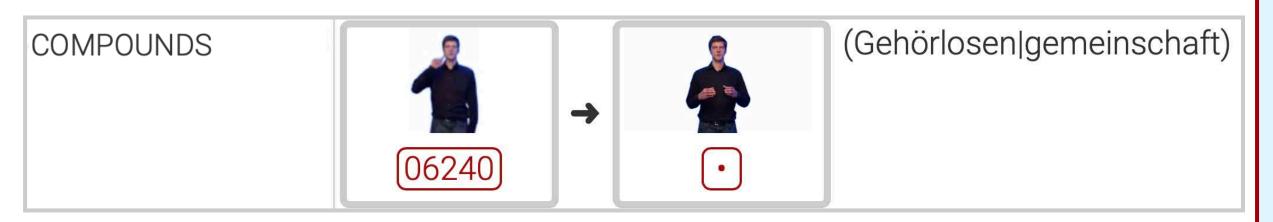| PHRASE #10	COMPONENTS	262 420 →
	MOUTHING	warm
	DEFINITION	atmosphere of solidarity and warmth within a group or community
	ENGLISH TRANSLATIONS	cordial; feeling of belonging together; to feel connected

Phrase #10 combines a lexical variant 262 or 420 'warm' with 440. The combined meaning is not transparent. The phrase also occurs in entries 262 and 420.

Loan Compounds and Compound-like Compositions
- Sign combinations that mirror the sequence and structure of German compounds; the German compound is usually simultaneously mouthed
- Individual signs of the combination strongly correspond to individual parts of the mouthing but several different variants may be used for each element
- Loan compounds are dominated by the German compound visible in the mouthing

- A selection of compounds and compound-like compositions found in the corpus are documented as run-ons in the entry.

COMPOUNDS 06240 → (Gehörlosen|gemeinschaft)

Sign 440 combines with sign 06240 'deaf', resulting in a compound(like) composition as 'deaf community'. The compound may be a loan or a parallel structure.

Reference

Bo Svensén 2009: A Handbook of Lexicography. The Theory and Practice of Dictionary-Making. Cambridge: Cambridge Univ. Pr.