Introducing the DW-DGS – The Digital Dictionary of DGS

Gabriele Langer, Anke Müller, Sabrina Wähl, Felicitas Otte, Lea Sepke, Thomas Hanke
Institute of German Sign Language and Communication of the Deaf, University of Hamburg, Germany
{gabriele.langer, sabrina.waehl, felicitas.otte, thomas.hanke}@uni-hamburg.de
ankemueller.am@posteo.de, lea.sepke@outlook.de

Abstract

This article describes the lexical resource DW-DGS – the first corpus-based digital dictionary of German Sign Language (DGS). Basic information is provided on dictionary type, context of compilation, sign representation in the product, metalanguage, dictionary content, information types displayed in entries, and dictionary structure. The article also provides an overview on data sources, methods, workflow procedures, and tools used in the lexicographic process. Challenges of making a corpus-based sign language dictionary and solutions developed for the DW-DGS are mentioned. The aim of this contribution is to provide an overview on the resource. It also serves as a starting point by referring to papers that describe the structures and procedures of the DW-DGS in more depth.

Keywords: sign language dictionary, lexical resource, German Sign Language (DGS), corpus-based lexicography

1. General Information

The full title of the online electronic dictionary of DGS described here is *Digitales Wörterbuch DGS (DW-DGS). Das korpusbasierte Wörterbuch DGS – Deutsch* [Digital Dictionary of DGS (DW-DGS). The corpus-based Dictionary DGS – German]. It is one of the products of the *DGS-Korpus project* (2009–2027).

The *DW-DGS* can be accessed at: https://dw-dgs.de. We refrained from including screenshots from the dictionary as figures and ask the reader to open the online dictionary for illustration. We recommend to look at entries 193, 366, 440 and 354 that cover most information types mentioned in this paper.

2. Dictionary Type

The *DW-DGS* is the first general corpus-based dictionary of German Sign Language (DGS). It is a descriptive dictionary produced in an academic context that focuses on the documentation of the general language of DGS. As a synchronic dictionary it targets at contemporary language – based on DGS as it was used at the time of data collection (2010–2012). The dictionary covers signs from all regions of Germany.

In the *DW-DGS*, established DGS signs are described from a primarily monolingual perspective on the basis of their uses in context as evidenced in the data of the *DGS Corpus*. The dictionary is corpus-based and largely, but not completely corpus-bound. Following the well-established corpus-based approach of modern lexicography, the results of corpus analyses for each lemma sign are summarised in the dictionary entries (cf. e.g. Atkins and Rundell, 2008; Sinclair, 2003). The metalanguage used for description is German.

In addition to the description of DGS signs from a monolingual perspective the dictionary also provides some bilingual features. Senses of DGS signs listed and described in the entries are matched to German translational equivalents. This enables using the *DW-DGS* in the function of a bilingual dictionary DGS→German. The German index provides access to the DGS entries via German words thus fulfilling the function of a bilingual dictionary German→DGS. The *DW-DGS* can therefore be described as monolingual dictionary of DGS with additional bilingual features, or as a bilingualised monolingual dictionary (cf. Hannay, 2003; Svensén, 2009).

As far as the medium and conditions of publication and use are concerned, the *DW-DGS* is an electronic online dictionary that can be accessed freely and free of charge on the internet. It includes video clips of signs and signed example sentences.

The *DW-DGS* is made for a wide audience of users including the user groups of L1 DGS signers, L2 learners of DGS, DGS teachers, DGS interpreters, linguists, and the interested public.

For a discussion on the dictionary type, languages in the *DW-DGS*, and user groups cf. Langer et al. (2018b, 2022) and Müller et al. (2022).

3. Data Sources

Information provided in the *DW-DGS* takes into consideration data from three sources: The main source used is the *DGS Corpus*. It is supplemented to a small extent by data elicited via the *DGS-Feedback* and through *SignHunter*.

Proceedings of the 11th Workshop on the Representation and Processing of Sign Languages, pages 194–203
25 May 2024.

3.1. DGS Corpus

The *DGS Corpus* has been designed explicitly with the aim to provide a basis for the first corpus-based dictionary of DGS. Its current size is more than 680.000 tokens (Feb 2024).

So far the dictionary is not only based on corpus data but also largely corpus-bound. We do not include signs or senses that are not evidenced in the corpus data. For reasons of reliability, information on meaning and usage are based on analyses of fluent signing in context. Entry information are abstractions from corpus evidence. For corpus analyses all *DGS Corpus* data available of a lemma sign are used. That includes data published in the *Public DGS Corpus* but also lemmatised unpublished data. The *DGS Corpus* data is stored, annotated and worked with in iLex. iLex is the lexical database and annotation environment that is used in the project for annotation, data curation and analyses. In iLex the data is matched to a hierarchy of type and subtype entries.

For further information on corpus design, elicitation tasks, data collection and corpus curation cf. Schulder et al. (2021). For the concept of types and subtypes and the type structure in the iLex database cf. Langer et al. (2018a). Langer et al. (2016b), especially the poster, includes an example illustrating the different type levels and their use in the iLex database. Type levels as displayed in the *Public DGS Corpus* are explained in Konrad et al. (2022). For more information on iLex cf. Hanke (2002).

3.2. DGS-Feedback

Some usage data for a small set of signs have been collected online from signers via the so-called *DGS-Feedback*. Participants were presented sign-meaning combinations and asked whether they used or knew these signs for these meanings. These data are used in addition to complement, clarify or solidify the results from corpus data.

More information on how the data collected by the *DGS-Feedback* is used in compiling entries for the dictionary cf. Wähl et al. (2018). For a description of the *DGS-Feedback*, the design of the questionnaires and the question types cf. Matthes et al. (2014). For technical aspects of the *DGS-Feedback* system cf. Berding and Hanke (2015).

3.3. SignHunter

SignHunter is a tool that was created and used to collect additional data from participants at deaf events. Participants are presented isolated stimulus items and can choose items for which they want to contribute and record their signs. *SignHunter* was used merely for concepts that were considered unambiguous. So far signs for city names and signs for social media names have been collected via *SignHunter*.

For more information on the data collection tool *SignHunter* cf. Hanke et al. (2020).

4. Representation of Signs

There is no established, widely known writing system for DGS that could be used to represent DGS in the dictionary. As we do not expect the occasional user to learn a notation system just to be able to consult the dictionary, we decided against using notations. Glosses were not an option either: They bear the risk of interference by the gloss word and conflict with the idea of representing signs as entities of their own, spoken-language independent visual nature. Instead, signs are represented either by recorded videos or by small visual elements called micons. A micon is a thumbnail movie displaying the form of the lemma sign combined with a unique entry number for quick identification and reference. Hovering over the micon's thumbnail sets the micon in motion, clicking on it plays a larger video of the sign in the movie display area, clicking on the number below the thumbnail opens the corresponding entry.[1]

In the entries, micons are used as sign representations for information types given in DGS such as synonyms, antonyms, collocational patterns and multi-sign expressions where they also serve as implicit cross-references. Outside the entries, micons are used to represent lemma signs in access structures.

For more detail on the rationale of a gloss-free dictionary and the use and function of micons as a means of lemma sign representation cf. Langer et al. (2018b, 2022, 2019) and Otte et al. (2022).

5. Metalanguage

Written German is not only one of the target languages of the dictionary, but it is also used for the dictionary definitions, descriptions, comments and subject categories in the entry, as well as category headings and other elements used for orientation or navigation, such as menu options and buttons.

Front matter information is also provided in written German. While a signed version is not yet complete, users of DGS find related information in a set of tutorials explaining the *DW-DGS* in DGS.

[1]The term 'micon' is derived from 'moving icon' and was first coined by Russel Sasnett (Brøndmo and Davenport, 1989). In this original use, 'micon' referred to the small video playing in loop on its own. We have adapted the term for our purposes to include the ID number as well.

The tutorials in DGS can be found at https://dgs-korpus.de/tutorials.html. On the rationale for using German as metalanguage cf. Langer et al. (2022) and Müller et al. (2022).

6. Content of the DW-DGS

6.1. Signs

The dictionary describes established manual signs of DGS. Only simplex signs are treated as lemma signs and are given entry status. Multi-sign expressions aka multi-word expressions (MWE) are not treated as lemma signs in their own right. They are to some extent included and appear within entries at different places, either on the sense level, as information addressed to a sense, or in the run-on section of an entry. For the time being there are no entries for productive forms i.e. classifier signs or classifier handshapes, nor for non-manual elements.

Signs and senses listed in the *DW-DGS* are largely restricted to what is evidenced in the *DGS Corpus*. Sign variants, that is lexical and phonological variants, are included. Lemma selection is guided by frequency.

6.2. Information Types

The entries of the *DW-DGS* contain several different kinds of information. The following information types relate to the lemma sign as a whole:

1. *Form*: information on form and form variants provided as studio recordings;
2. *Kommentar*: comments on aspects of form, usage and other additional information on the sign;
3. *Beleglage*: rough indication of frequency of the sign in the *DGS Corpus*;
4. *Grammatik*: grammatical label or comment;
5. *Regional*: comment on regional distribution, including distributional maps;
6. *Bedeutung*: information on meaning and use: list of senses in the form of signposts [2];
7. *Zusammensetzungen*: compound-like constructions containing the sign;
8. *Verwandt*/*Formgleich*/*Formähnlich*: cross references to related signs and signs of the same form or a similar form, and
9. *Konkordanz*: concordance view of tokens of the sign in the *Public DGS Corpus*.

The following information types relate to a particular sense in the senses' section:

10. rough indication on the meaning of the sense (*Signpost*);
11. *Form* (only provided for phrases and multi-sign names);
12. *Mundbild*: selection of typical mouthings or information on mouth gesture used with the sign and a studio recording of the sign with a typical mouthing or mouth gesture;
13. *Erklärung*: explanation of the sense, the so-called dictionary definition;
14. *Deutsch*: German translational equivalents, sometimes with disambiguation information or diasystematic label;
15. *Anmerkung*: additional information on usage;
16. *Grammatik*: grammatical information specific to the sense;
17. *Beispiele*: authentic examples illustrating the sense, each with a clip of the original *DGS Corpus* recording, a German translation and a short context, and with direct links to its original location in the two portals of the *Public DGS Corpus* (*MY DGS* and *MY DGS – annotated*);
18. *Bedeutungsgleich*: synonym and near-synonym signs, sometimes with a clickable thumbnail map that displays the regional distribution of a set of coexisting lexical regional variants;
19. *Entgegengesetzt*: antonym signs of opposite or complementary meaning sometimes with a clickable thumbnail map that displays the regional distribution of a set of coexisting lexical regional variants;
20. *Häufige Kombinationen*: collocational patterns and semantic preference patterns;
21. *Zusammensetzungen*: compound-like constructions that can be related to this particular sense;
22. *Regional*: comment on the regional distribution of this sense or a group of senses with a link to a corresponding distribution map;
23. *Sachgruppen*: subject areas that this sense is assigned to.

[2] Within the list of senses some MWE are listed under the categories phrase (*Phrase*) and multi-sign name (*Mehrteiliger Name*).

Not all types of information are given in each entry or for each sense. Information is provided only when relevant and available.

Langer et al. (2022) provides further details on the information types mentioned here.

6.3. Types of Entries

Entries in the *DW-DGS* differ with respect to their analytic and descriptive depth. This is partly due to the varying amount and quality of data available in the corpus for each lemma sign and partly due to issues of time and resources.

While lexicographers explored what could be done with corpus data at hand for a larger number of entries, it was not possible to invest the same amount of time and labour in the preparation of all entries in the same way. As a consequence the team opted for a mixture of more and less elaborated entries. The dictionary contains elaborated entries and shorter entries with less fine-grained sense distinctions and less information. Also, there are entries completely edited by lexicographers and entries that have been partly edited starting with automatically compiled data.

A third kind of entries in the *DW-DGS* are automatically compiled entries. For these entries only minimal editing steps such as lemma establishment were done manually. Information provided in automatically compiled entries includes senses inferred from cross-references originating at manually edited entries as well as rough meaning indications, i.e. German equivalents, already prepared in iLex for subtypes of the *DGS Corpus* types list. Automatically compiled entries have not yet received a full lexicographic treatment. Such entries are not included in the DGS index but can be accessed from cross-references addressing them in edited entries and through their listing in the German index. Automatically compiled entries can be identified by their micon appearance (red number on white background as opposed to the white number on red background shown for edited and partly edited entries) and by the heading *(Automatisch generierter Vorabeintrag)* at the top of the entry page.

For more information on different entry types in the *DW-DGS* cf. Wähl et al. (2022).

7. Dictionary Structure

7.1. Navigation: Menu Bar

The menu bar at the top of each *DW-DGS* web page enables the user to choose which part of the dictionary they want to visit. The default page is the body of entries (option: *DGS*) which is displayed by default when opening the dictionary URL.

7.2. Front Matter (*Intro*)

The front matter of the *DW-DGS* contains an introduction including information on dictionary use (user's guide), background information on the data used, the lexicographic process, maps, and relevant object language information for DGS.

7.3. Back Matter (*Karten*)

In the back matter, the users find a number of maps including geographical distribution maps of coexisting regional signs belonging to specific semantic sets, such as signs for the days of the week or colors, and interactive geographical maps with city and country name signs.

7.4. Access Structures

While in print dictionaries there is one primary sort key determining the order of entries in the main part of the dictionary and several indexes on secondary sort keys, the *DW-DGS*, like many electronic dictionaries, has individual pages for each of the entries and several indexes providing access to the individual pages.

For more information on the access structures available in the *DW-DGS* cf. Langer et al. (2022).

7.4.1. Macrostructure (*DGS*)

The main and most important index of the *DW-DGS* shows the body of entries. Each entry is represented by a micon. The macrostructure consists of a table of all micons. The user can choose between several options of ordering the entries represented by micons: by entry number (*Nummer*), which is also the default, by handshape (*Handform*), by number of hands (*Händigkeit*), or by place of articulation (*Lokation*). The secondary sort key is the height of the place of articulation from high to low. Where the variants of an entry differ with respect to their values for the current sort keys, the variants are shown separately, i.e. the micons then represent individual variants and are thus marked with a .1, .2, etc. appended to the entry number. This ordering allows for a very rough search by form.

7.4.2. German Index (*Deutsch*)

The German index is an additional access way to the information provided in the sign entries and supports searches for signs through German words. It consists basically of a table with a listing of German words in the first column. In many cases the German word is disambiguated by a context in the second column. Micons represent the corresponding sign senses in the third column. In some cases of high regional lexical variation an additional thumbnail map is displayed in the third column as a cross

reference to the cluster map visualising regional distribution of coexisting lexical variants.

The German index is generated from the German translational equivalents provided in the entries.

The German words in the German index do not receive the same in-depth attention and treatment as lemma signs of their own right as they would in a fully bilingual dictionary. No missing words or word senses are added.

The German index provides dictionary-external links to corresponding entries in the corpus-based German Dictionary *DWDS* (Berlin-brandenburgische Akademie der Wissenschaften) where additional information on the German words can be looked up quickly if desired. This compensates somewhat for the scarceness of information provided for German equivalents.

7.4.3. Subject Area Index (*Sachgruppen)*

In the entries, each sense is matched to up to three subject areas in which the the sign-sense combination is then listed. The subject area index is a topic-specific way to access the signs contained the *DW-DGS*. In a table it lists subject areas together with the signposts and micons of the senses allocated to them.

7.4.4. Graph (*Graph)*

The graph is a visual structure that provides a non-text-based access way to the dictionary. Entries are depicted as dots. A clickable micon appears when the cursor hovers over a dot. The dots are connected by color-coded lines that represent different relation categories between entries, that is synonyms (*Bedeutungsgleich*), antonyms (*Entgegengesetzt*), collocations (*Häufige Kombinationen*), compoundlike constructions (*Zusammensetzungen*), parts of MWE (*Bestandteile*), signs having the same form (*Formgleich*), signs having a similar form (*Formähnlich*), and related signs (*Verwandt*). The user can modify the graph to show only certain kinds of connections by unclicking all other checkboxes. The graph is a tool for playful exploration of the dictionary.

For more information on visual access to the dictionary by the graph cf. Langer et al. (2022), a short description can be found in Müller et al. (2022).

7.5. Microstructure: Entries

Each entry has its own web page with a unique entry number for identification at the top, a video display area, and a table containing the entry information. In the first column of the table the category labels of the information types are given while the second column contains the information provided. The head section shows information addressed to the whole sign. It is followed by the list of senses. At the bottom run-on information such as MWE and form-related cross references to other signs are given.

The middle part contains the list of senses. In its collapsed state it is presented as a list of signposts that hint on the senses' meanings. Each sense row is numbered and can be expanded to reveal all information addressed to the sense. When expanded, category labels for the information types addressed to the sense are displayed in the second column while the corresponding information is provided in an additional column to the right. Information given in DGS is either displayed as micons or can be viewed as movie in the video display area by clicking on the button with the play-symbol.

7.6. Mediostructure: Cross Referencing

In the *DW-DGS* all cross-references to lemma signs, variant forms, or senses are realized as micons. The thumbnail micon figure represents the lemma sign in the form of either the first variant as default or, when relevant, a different variant of the lemma sign. The specific address is expressed by the number of the micon: Cross-referenced lemma signs (i.e. whole entries) appear with the entry number only (e.g. 144), micons for cross-referenced variants with entry number followed by a point and the number of the variant form (e.g. 144.2), cross-referenced senses with entry number followed by a hash and the sense number(s) (e.g. 144#2). Micons function as implicit cross references as there is no special reference marker. Entry-internal and entry-external cross references are not distinguished visually.

Hyperlinks to entries in the German index or in the subject area index use the written German words as labels set in blue text color.

The dictionary contains two kinds of maps, one showing the regional distribution of a single lemma sign and the second showing several coexisting lexical, mostly regional variants in contrast to each other (cluster maps). For the first kind, *Karte* (map) hyperlinks lead to extra map pages. Cross references to cluster maps are realised by clickable thumbnails of the map.

8. Links to other Resources

The *DW-DGS* is an online resource that makes use of the possibility to include cross references that directly link to resources outside the dictionary. For more detail on linking to and fro the *DW-DGS* cf. Müller et al. (2020).

8.1. Linking to the Public DGS Corpus

At two different places in the *DW-DGS* external links to the *Public DGS Corpus* are provided.

Example sentences link via buttons below the video display area to their location in the two portals of the *Public DGS Corpus*. These buttons are not shown when examples are taken from unpublished parts of the *DGS Corpus*.

At the very bottom of each entry the button *Konkordanz* opens a page with a concordance view of all tokens in the annotated Public DGS Corpus (*MY DGS – annotated*) that are realisations of the lemma sign of the respective *DW-DGS* entry. From the concordance view one can jump to the corresponding types list entry of *MY DGS – annotated* by clicking on a gloss or into the transcript by clicking on the transcript's name in the upper left side of a concordance line.

8.2. Linking to the DWDS

In the German index, links to the German online dictionary called *DWDS* are provided whenever a match could be found for a listed German word or word sense. The links are realised in form of clickable blue and white *DWDS* logos.

9. Method and Workflow

For the lexicographic process from data analysis to the finished dictionary entry, we adapted the lexicographic principles and steps of corpus-based lexicography as described in Atkins and Rundell (2008) for sign language lexicography.

For a short description of the steps in the lexicographic process cf. Langer (2021) and Langer et al. (2018a).

9.1. Dictionary Writing System

We use a FileMaker database as our dictionary writing system (DWS). Filemaker is a low-code program in which the user is able to configure and adjust the user interface without programmer's support. Some information from iLex such as type glosses and HamNoSys notations is imported for direct display into the DWS. Other entities are entered with only their iLex ids for reference, e.g. types used for cross references and tags needed for authentic example management. These entities as a well as SQL queries can be opened directly in iLex via scripts stored in the FileMaker database. In the DWS, types and subtypes are grouped for lemma establishment, pre-lexicographic information is stored, and entry information is prepared and edited for publication.

9.2. Lemma Selection

The first step of the lexicographic process is lemma selection. For the *DW-DGS* this basically driven by frequency. The general threshold for inclusion of a corpus type candidate into the dictionary is that it contains at least one subtype with at least 25 tokens from a number of different signers. In certain cases we work with less than 25 tokens, for example when we are dealing with lexical variants that are used only in certain geographic areas and are part of a group of several coexisting regional signs for one concept, or when the sign in question is part of a semantic set that would be missing one element just because of low token numbers. Tokens of phonological variants may be added up to meet the threshold while so-called non-tokens should be excluded.

Lemma selection is described in more detail in Wähl et al. (2022). A description and discussion of non-tokens can be found in Langer et al. (2016a).

9.3. Establishment of Lemma Signs

For each selected lemma sign candidate it has to be determined which subset of tokens, i.e. which types and subtypes are apt to constitute the data to be described in this sign's dictionary entry. During this step also other variants, related or similar types have to checked for possible inclusion. Inspired by Svensén (2009, p. 94), we call this establishment of lemma signs to distinguish it from lemmatisation in annotation.

For DGS the establishment of lemma signs is much less straightforward than for a well researched spoken language with a long codified written tradition. In DGS we find a high variation in form, iconic modifications of sign forms, and a somewhat flexible combination of signs with mouthings that contribute to the semantics of the signs. Often the lexicographers are confronted with a large number of similar signs with only partly overlapping meanings. This makes lemma establishment in DGS a rather challenging task. The process requires a number of different aspects, principles and criteria that have to be taken in consideration and weighed against each other. While in principle lemma selection and the establishment of lemma signs are two separate steps, in practice they are mutually dependent and thus done at the same time.

The lemma establishment rules and principles used for the *DW-DGS* with illustrating examples can be found in Langer et al. (2020). Hanke et al. (2023) describes an example where the regional distribution of subtypes is considered in the decision making for the establishment of lemma signs. For sign languages issues of lemma sign establishment have also been described and discussed by Johnston and Schembri (1999); Kristoffersen and

Troelsgård (2010); Fenlon et al. (2015).

9.4. Compiling Entries

Compiling entries is a complex task in which analysing available data, abstracting, summarising and describing the results while preparing the entries for publication goes back and forth. Corpus data analyses during this task include a look at form variation and sign forms in fluent signing, regional distribution, distribution across age groups, and frequent neighbours (collocations). The citation form and variant forms to be included are determined. The central and most time-intensive task is Word Sense Discrimination (WSD). Once the senses of of a lemma sign have been determined and described in the DWS all other information addressed to a sense can be entered and prepared for publication.

The various steps during compiling entries of the *DW-DGS* are described in Langer (2021); Langer et al. (2018a).

9.4.1. Word Sense Discrimination (WSD)

WSD encompasses identifying meanings and uses of a lemma sign and describing them as a list of sign senses. For this task a considerable number of tokens are viewed in context, that is, the recorded movies are viewed, alongside with annotations and translations. Lists of frequent neighbours in the corpus can help to identify different uses. Results of recurring similar uses are summarised and entered as senses in the DWS database. Each sense is described by a German explanation, the so-called dictionary definition. Corpus tokens contributing to the evidence for a sense are tagged in iLex and suitable candidates for dictionary examples are referenced in the DWS for further use. In a second step the proto-senses are reviewed, lumped when necessary, and marked for production or exclusion.

9.4.2. Editing the Entry

Usage examples are selected from the candidates to illustrate meanings and typical uses of a sign in context. They are prepared for publication and receive a short context and the translation is adapted for use in the dictionary. Synonymous or antonymous signs are included in the description of a sense if available in order to provide language-internal hints on the sign's sense. Further information that is part of a sign sense's section is prepared and entered into the DWS, including typical mouthings, sign combinations such as collocations and semantic preferences, German translational equivalents, subject areas, and information on usage and distribution across regions and age groups. Maps showing regional distribution are prepared and marked for publication when subtypes display a noticeable regional distribution. When necessary, disambiguating contexts are entered for translational equivalents. Equivalents that are less useful in a reversal of the search direction from DGS→German to German→DGS are marked for exclusion from the German index. Each sense also is provided a signpost for the meanings overview in collapsed entries.

For the extraction of a frequent neighbour lists from the DGS corpus and its use in analyses and inclusion in the DW-DGS cf. Langer (2021); Langer et al. (2018a); Langer and Schulder (2020). On the selection and preparation of authentic examples from the corpus cf. Langer et al. (2018b). More information on the use of maps for exploration in analyses and an example analysis of regional clusters is found in Hanke et al. (2023).

9.5. Production

9.5.1. Studio Recordings

Representing DGS in the dictionary requires studio recordings of the respective signs. The videos are needed to produce the still and the animated figure of the micon as well as to be played in the video display area to show the sign's form in isolation. At the level of the lemma sign, signs are recorded without mouthing to serve as information on variant forms. For each sense, an individual recording is made showing the sign with a typical mouthing or mouth gesture.

A list of required movies is generated from the DWS, matched against the list of movies already available in iLex from previous recordings, and provided as a script list for the studio recording. The required signs are performed by deaf models in our video studio. In total, seven different cameras from four different angles are used. In post-production movies are converted and integrated into iLex. Then the video material is annotated by student assistants and checked by deaf team members for whether the signs were correctly executed by the signing model. After that student assistants choose one of the frames of the video as the thumbnail for the micons.

9.5.2. Production of the Dictionary

For the production of an updated dictionary version relevant data for the prepared entries are exported from the DWS and converted into a json file. The production scripts generate maps, video clips of studio recordings and example sentence videos as well as the micons. Production data from the json file and data from iLex are combined to generate the html pages of the dictionary for the manually edited

entries as well as for the automatically compiled entries.

On technical aspects of generating maps from iLex data cf. Hanke (2018).

10. Highlights of the DW-DGS

The *DW-DGS* contains several interesting and new information types with regard to sign language dictionaries. Here only few highlights can be mentioned briefly.

One special trait of the *DW-DGS* are the authentic examples taken from the original recorded data. They illustrate and reinforce the more general and abstract sense descriptions and contribute an element of liveliness. In addition, the example movies grant visibility of the language community through its contributing members.

Through the corpus it has become possible to use statistics to identify frequent neighbours in DGS. To our knowledge it is the first time that collocation patterns were used for WSD of a sign language and are included as information in a sign language dictionary.

Maps showing the regional distribution for individual signs or several coexisting regional signs (cluster maps) are very attractive for many deaf users. These maps are directly generated from corpus data and participant metadata and would not be possible without data from the *DGS Corpus*.

The last information type provided in the *DW-DGS* that we want to highlight here are the cross references to sign synonyms, near synonyms and antonyms of a sense. They serve as monolingual explanatory elements. A robust and proper distinction and description of sign senses is a good basis and in our view almost a prerequisite for discovering and displaying information on synonym and antonym relations between signs.

11. The DW-DGS in Numbers

The *DW-DGS* contains 1876 entries: 802 fully edited, 272 partly edited entries and 802 automatically compiled entries (cf. section *6.3. Types of Entries* for more details on entry types). The 1074 edited entries contain 2436 senses with 3416 authentic examples and 6377 German equivalents. For 581 senses sense-related distribution maps are provided. The dictionary includes 50 distribution maps for sets of coexisting regional lexical variants. Additional 1290 cross-references between synonyms and near-synonyms and 726 cross-references between antonyms are provided between senses. The edited entries include 416 collocational patterns and 141 compound-like patterns. There are 2679 manually selected cross-references to signs of similar or same form between entries. (Numbers date from end of March 2024.)

12. SL-specific Challenges

There are many challenges specific to corpus-based sign language lexicography. A high variation of sign forms with many partly overlapping meanings in a corpus of still limited size make lemma sign establishment complicated.

The lack of a written direct representation of the signing in the corpus along with only limited tools for corpus annotation and analysis makes working with signed corpus data a very time-consuming task as lexicographers cannot skim-read through samples but have to resort to watching the original video data one by one and sometimes several times to compare.

For the design and structure of a dictionary, the lack of a writing system and orthography for signed text results in the issues of sign representation, entry ordering (macrostructure), search for sign form, and the choice of metalanguage in the dictionary.

For a short discussion of some of these challenges cf. Müller et al. (2022).

13. Outlook

Corpus annotation is on-going. This enables us to expand existing entries as well as to create new ones. The *DW-DGS* is updated several times a year.

For searching a specific sign by its form, the *DW-DGS* currently offers different sort orders for the body of entries. There, the sign looked for then needs to be identified by browsing through the corresponding subsection. This becomes tedious if such a subsection contains too many items or the user is unsure about the location used in the sign (the secondary order criterion). Filtering facilities as implemented e.g. by the *ODT* (Centre for Sign Language, 2008–ongoing) and the *GaLex* (Konrad et al., 2010) might be of some help here, but this approach needs to be tailored to the size of entries in order to provide result sets small enough that browsing after filtering still appears natural to the user. So such a functionality can only be implemented now that the dictionary growth is expected to be much slower than in the past years.

14. Acknowledgements

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 Academies of Sciences and Humanities.

15. Bibliographical References

B. T. Sue Atkins and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford University Press, Oxford; New York.

Sven Berding and Thomas Hanke. 2015. Documentation of the Feedback-System and its integration into iLex. Project Note AP04-2015-01, Universität Hamburg, DGS-Korpus project, IDGS, Universität Hamburg.

Hans Peter Brøndmo and Glorianna Davenport. 1989. Creating and viewing the elastic charles – A hypermedia journal. In *Hypertext: State of the Art. Papers presented at the Hypertext 2 conference, York, 1989*, pages 43–51. Intellect Ltd., Oxford.

Jordan Fenlon, Kearsy Cormier, and Adam Schembri. 2015. Building BSL SignBank: The Lemma Dilemma Revisited. *International Journal of Lexicography*, 28(2):169–206.

Thomas Hanke. 2002. iLex - A Tool for Sign Language Lexicography and Corpus Analysis. In *3rd International Conference on Language Resources and Evaluation*, pages 923–926, Las Palmas, Canary Islands, Spain. European Language Resources Association (ELRA).

Thomas Hanke. 2018. Using d3js to visualise data in iLex. Project Note AP04-2016-01, Universität Hamburg, DGS-Korpus project, IDGS, Universität Hamburg.

Thomas Hanke, Elena Jahn, Sabrina Wähl, Oliver Böse, and Lutz König. 2020. SignHunter – A Sign Elicitation Tool Suitable for Deaf Events. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, pages 83–88, Marseille, France. European Language Resources Association (ELRA).

Thomas Hanke, Reiner Konrad, and Gabriele Langer. 2023. Exploring regional variation in the DGS Corpus. In Ella Wehrmeyer, editor, *Advances in Sign Language Corpus Linguistics*, number 108 in Studies in Corpus Linguistics, pages 192–218. Benjamins, Amsterdam.

Mike Hannay. 2003. 3.1 Types of bilingual dictionaries. In Piet van Sterkenburg, editor, *Terminology and Lexicography Research and Practice*, volume 6, pages 145–153. Benjamins, Amsterdam.

Trevor Johnston and Adam C. Schembri. 1999. On Defining Lexeme in a Signed Language. *Sign Language & Linguistics*, 2(2):115–185.

Reiner Konrad, Thomas Hanke, Gabriele Langer, Susanne König, Lutz König, Rie Nishio, and Anja Regen. 2022. Public DGS Corpus: Annotation conventions. Project Note AP03-2018-01, Universität Hamburg, DGS-Korpus project, IDGS, Universität Hamburg.

Jette Hedegaard Kristoffersen and Thomas Troelsgård. 2010. Making a Dictionary without words: Lemmatization problems in a sign language dictionary. In Sylviane Granger, editor, *2010 eLexicography in the 21st Century: New Challenges, New Applications. Proceedings of eLex 2009, Louvain-la-Neuve, 22-24 October 2009*, number 7 in Cahiers Du Cental, pages 165–172. Presses Univ. de Louvain, Louvain.

Gabriele Langer. 2021. Vorgehen bei der Analyse für die Artikelschreibung (Wörterbuch). Arbeitspapier AP10-2016-02, Universität Hamburg, DGS-Korpus project, IDGS, Universität Hamburg.

Gabriele Langer, Thomas Hanke, Reiner Konrad, and Susanne König. 2016a. "Non-tokens": When Tokens Should not Count as Evidence of Sign Use. In *Proceedings of the LREC2016 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining*, pages 137–142, Portorož, Slovenia. European Language Resources Association (ELRA).

Gabriele Langer, Anke Müller, Felicitas Otte, and Sabrina Wähl. 2022. Information Types and Use Cases. Project Note AP10-2021-02, Universität Hamburg, DGS-Korpus project, IDGS, Universität Hamburg.

Gabriele Langer, Anke Müller, and Sabrina Wähl. 2018a. Queries and Views in iLex to Support Corpus-based Lexicographic Work on German Sign Language (DGS). In *Proceedings of the LREC2018 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community*, pages 107–114, Miyazaki, Japan. European Language Resources Association (ELRA).

Gabriele Langer, Anke Müller, Sabrina Wähl, and Julian Bleicken. 2018b. Authentic Examples in a Corpus-Based Sign Language Dictionary – Why and How. In *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts (EURALEX 2018)*, pages

483–497, Ljubljana, Slovenia. Ljubljana University Press.

Gabriele Langer, Anke Müller, Sabrina Wähl, and Thomas Hanke. 2019. The DGS-Korpus approach to including frequent sign combinations in a corpus-based electronic sign language dictionary.

Gabriele Langer, Anke Müller, Sabrina Wähl, Susanne König, Thomas Hanke, and Reiner Konrad. 2020. Lemmatisierungsregeln (Vorläufige Wörterbuch-Einträge). Arbeitspapier AP10-2016-01, Universität Hamburg, DGS-Korpus project, IDGS, Universität Hamburg.

Gabriele Langer and Marc Schulder. 2020. Collocations in Sign Language Lexicography: Towards Semantic Abstractions for Word Sense Discrimination. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, pages 127–134, Marseille, France. European Language Resources Association (ELRA).

Gabriele Langer, Thomas Troelsgård, Jette Kristoffersen, Reiner Konrad, Thomas Hanke, and Susanne König. 2016b. Designing a Lexical Database for a Combined Use of Corpus Annotation and Dictionary Editing. In *Proceedings of the LREC2016 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining*, pages 143–152, Portorož, Slovenia. European Language Resources Association (ELRA).

Silke Matthes, Thomas Hanke, Susanne König, and Gabriele Langer. 2014. Entwicklung des DGS Feedback-Systems. Arbeitspapier AP07-2014-01, Universität Hamburg, DGS-Korpus project, IDGS, Universität Hamburg.

Anke Müller, Thomas Hanke, Reiner Konrad, Gabriele Langer, and Sabrina Wähl. 2020. From Dictionary to Corpus and Back Again – Linking Heterogeneous Language Resources for DGS. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, pages 157–164, Marseille, France. European Language Resources Association (ELRA).

Anke Müller, Gabriele Langer, Felicitas Otte, and Sabrina Wähl. 2022. Creating a dictionary of a signed minority language. A bilingualized monolingual dictionary of German Sign Language. In *Proceedings of the XX EURALEX International Congress: Lexicography in Global Contexts (EURALEX 2022)*, pages 635–648, Mannheim, Germany. IDS-Verlag.

Felicitas Otte, Anke Müller, Gabriele Langer, Sabrina Wähl, and Thomas Hanke. 2022. Sign representation in the DW-DGS. Project Note AP11-2021-01, Universität Hamburg, DGS-Korpus project, IDGS, Universität Hamburg.

Marc Schulder, Dolly Blanck, Thomas Hanke, Ilona Hofmann, Sung-Eun Hong, Olga Jeziorski, Lutz König, Susanne König, Reiner Konrad, Gabriele Langer, Rie Nishio, and Christian Rathmann. 2021. Data statement for the Public DGS Corpus. Project Note AP06-2020-01, Universität Hamburg, DGS-Korpus project, IDGS, Universität Hamburg.

John Sinclair. 2003. Corpora for lexicography. In Piet van Sterkenburg, editor, *A Practical Guide to Lexicography*. Benjamins, Amsterdam.

Bo Svensén. 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge University Press, New York.

Sabrina Wähl, Gabriele Langer, and Anke Müller. 2018. Hand in Hand - Using Data from an Online Survey System to Support Lexicographic Work. In *Proceedings of the LREC2018 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community*, pages 199–206, Miyazaki, Japan. European Language Resources Association (ELRA).

Sabrina Wähl, Gabriele Langer, Felicitas Otte, Anke Müller, Thomas Hanke, and Reiner Konrad. 2022. Lemma Selection and Entry Types. Project Note AP10-2021-01, Universität Hamburg, DGS-Korpus project, IDGS, Universität Hamburg.

16. Language Resource References

Berlin-brandenburgische Akademie der Wissenschaften. DWDS – Digitales Wörterbuch der deutschen Sprache. Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart.

Centre for Sign Language. 2008–ongoing. Ordbog over Dansk Tegnsprog. [Dictionary of Danish Sign Language.].

Reiner Konrad, Gabriele Langer, Susanne König, Thomas Hanke, and Christian Rathmann. 2010. Fachgebärdenlexikon Gärtnerei und Landschaftbau.