

Corpus à la carte – Improving Access to the Public DGS Corpus

Reiner Konrad , Thomas Hanke , Amy Isard , Marc Schulder ,
Lutz König , Julian Bleicken , Oliver Böse 

Institute of German Sign Language and Communication of the Deaf

University of Hamburg, Germany

{reiner.konrad, thomas.hanke, amy.isard, marc.schulder,
lutz.koenig, julian.bleicken, oliver.boese}@uni-hamburg.de

Abstract

This article presents the fourth release of the Public DGS Corpus, a large corpus of German Sign Language (DGS). Since its first release in 2018, the Public DGS Corpus has provided its content through multiple portals to meet the needs of different user groups. Having started with a community portal and a research portal for general data access, the ANNIS portal for dynamic web-based exploration of the corpus was added in 2022. With this latest release, a fourth portal is added to allow sign language linguists to access the public corpus directly through the annotation software iLex. Furthermore, search capabilities and interconnectedness between the portals are strongly improved, allowing users to move between portals to combine their strengths. Additional improvements to the corpus include additional recordings, new pose information models, improved HamNoSys, enhanced type information and web interface revisions.

Keywords: German Sign Language (DGS), Linguistic Resource, Corpus, Resource Extension

1. Introduction

This paper presents the fourth release of the Public DGS Corpus, introducing new features and content as well as a new portal, *MY DGS – iLex*. It is a follow-up of Jahn et al. (2018) and Hanke et al. (2020), which described the previous releases of the corpus. A special focus is given to demonstrating how interconnectedness between the different corpus portals helps improve access to the data of the Public DGS Corpus. The paper also introduces the upcoming second data collection phase of the DGS Corpus, in which new primary materials will be recorded.

The paper is structured as follows: Section 2 gives a brief description of the DGS-Korpus project, the upcoming second data collection, and the history of releasing the Public DGS Corpus through a set of portals optimised for different use cases. The new features and content of release 4, including the new portal *MY DGS – iLex*, are presented in Section 3. Section 4 describes the connections between the portals, and Section 5 works through two use case examples which illustrate how connections between the portals can be used to combine the different strengths of each portal to match the user's needs, making the Public DGS Corpus a "corpus à la carte".

2. The DGS Corpus

The DGS-Korpus project (2009–2027) is a long-term research project to create a reference corpus of German Sign Language (DGS; *Deutsche Gebärdensprache*) (Prillwitz et al., 2008). Building on the DGS Corpus, the project has also produced a corpus-based dictionary, *DW-DGS* (Langer et al., 2024) and the Public DGS Corpus dataset which is the focus of this article. The purpose of the corpus is to be both a resource for linguistic research and a record of deaf heritage in Germany.

densprache) (Prillwitz et al., 2008). Building on the DGS Corpus, the project has also produced a corpus-based dictionary, *DW-DGS* (Langer et al., 2024) and the Public DGS Corpus dataset which is the focus of this article. The purpose of the corpus is to be both a resource for linguistic research and a record of deaf heritage in Germany.

2.1. Data Collection Phases

The majority of data for the DGS Corpus was gathered during its first data collection phase (2010–2012), during which dyadic conversations between 330 participants from thirteen regions of Germany and four different age groups were recorded, resulting in 1150 hours of recordings, containing 560 hours of semi-spontaneous DGS signing.

A second data collection is scheduled for 2024–2025 to add a fifth cohort of 46 participants, most of whom will be aged 18–32, i.e. people who have come of age since the first collection phase. Performing this second collection with a younger cohort allows the corpus to cover further relevant developments affecting the German deaf community, such as changes in educational policy (progressing from bilingual pilot projects to integration and then to inclusivity), changes in information technology (use of smart phones, social media, video telephony), medical advances in the use of cochlear implants, demographic changes, international mobility and language contact.

The second data collection will follow the design of the first collection, with a number of updates to account for changes in technology and participant

background. New camera equipment will be used, increasing video resolution from 2K to 6K. The collection formats will stay the same, but some tasks and stimuli are updated. For example, the list of historical events has been changed to match the time frame actively experienced by participants, a new discussion topic regarding social media has been introduced, and various images have been switched out to reference more recent politicians, celebrities and events as well as contemporary technology hardware (e.g. replacing images of CRT monitors with flatscreens).

2.2. The Public DGS Corpus: Thinking with Portals

A part of the DGS Corpus was selected for inclusion in a fully annotated publicly available dataset, the Public DGS Corpus.¹ The Public DGS Corpus was initially released in 2018 (Jahn et al., 2018) and extended in content and features through subsequent releases in 2019 and 2020 (Hanke et al., 2020). To accommodate the needs of different user groups (see Jahn et al., 2018), multiple web portals were created:

MY DGS is a community portal for the deaf community, presenting the corpus as a heritage resource, with a focus on easily finding interesting conversations about various aspects of deaf culture and life experience. To accommodate language learners and others interested in deaf culture, optional German subtitles are provided.

MY DGS – annotated targets linguistic researchers, providing all recordings with full annotations and translations in German and English, available for download and display through an online viewer. It also provides machine-readable metadata files and pose information for computational processing. In addition to the recordings of the community portal, some further recordings are provided that cover tasks with purely linguistic, rather than cultural, value, such as retellings of stories.

MY DGS – ANNIS has been available as a third portal since 2022, as described in Isard and Konrad (2022). ANNIS is a web-based corpus search tool which allows users to search and visualise corpus data (Krause and Zeldes, 2016). The portal integrates the Public DGS Corpus into an ANNIS instance to allow dynamic exploration of corpus contents without the need for installing annotation software.

MY DGS – iLex is a new portal introduced with release 4. It provides the content of the Public DGS Corpus as a relational database that can be accessed via iLex (Hanke, 2002), a tool for lexicographic and corpus linguistic research.

¹For a discussion of why only a part of the DGS Corpus is made public, see Schulder and Hanke (2022).

3. Changes in Release 4

In this section we describe the new features, functions and kinds of information made available by release 4 of the Public DGS Corpus. Section 3.1 describes the new corpus transcripts introduced by the release, which are added to all portals. The remaining descriptions are grouped by portal, starting with updates to *MY DGS* (Section 3.2) and *MY DGS – annotated* (Section 3.3), moving to how release 4 is integrated into *MY DGS – ANNIS* (Section 3.4) and a description of the new portal *MY DGS – iLex* (Section 3.5).

3.1. New transcripts

Release 4 introduces one hour of additional material to the Public DGS Corpus, bringing its full size to 52.4 hours. *MY DGS – ANNIS*, which excludes 2.4 hours of videos which have no annotations (see Isard and Konrad, 2022), grows from 49 to 50 hours.

The new material provides four retellings (7 minutes) and 18 process descriptions (53 minutes). The retellings are fully translated and lemmatised. They include two pear story retellings (Chafe, 1980) and two retellings of the broadcast “The Domestic Aid” (Sehen statt Hören, 2006). The new process descriptions are provided with translations, but without lemmatisation. They cover processes such as preparing a meal, baking a cake, or mending a puncture.

MY DGS does not cover certain tasks that are considered mainly of interest to linguistic research. For this reason, the new retelling recordings are omitted. However, for release 4 it was decided that process descriptions should be part of *MY DGS*, so the 18 new process descriptions are added, as well as the 13 process descriptions (42 minutes) that had been added to *MY DGS – annotated*. This brings the total size of *MY DGS* to 51.3 hours.

3.2. Changes to MY DGS

Release 4 of *MY DGS* offers enhanced search functionality to make content discovery even easier. In addition to filter options for region, age group, conversation format, and topics of conversation, a new full-text search on the German translations is introduced.

This search narrows down the selection of videos to those containing the searched text in their subtitles. When a specific video is selected, a list of all matching subtitle lines is shown, including the timecode at which it appears in the video and a button to start the video at that timecode. An example of this is shown in Figure 1. Text search also supports the wildcard symbol “*”, so as a side effect users can generate a full text transcript of the German

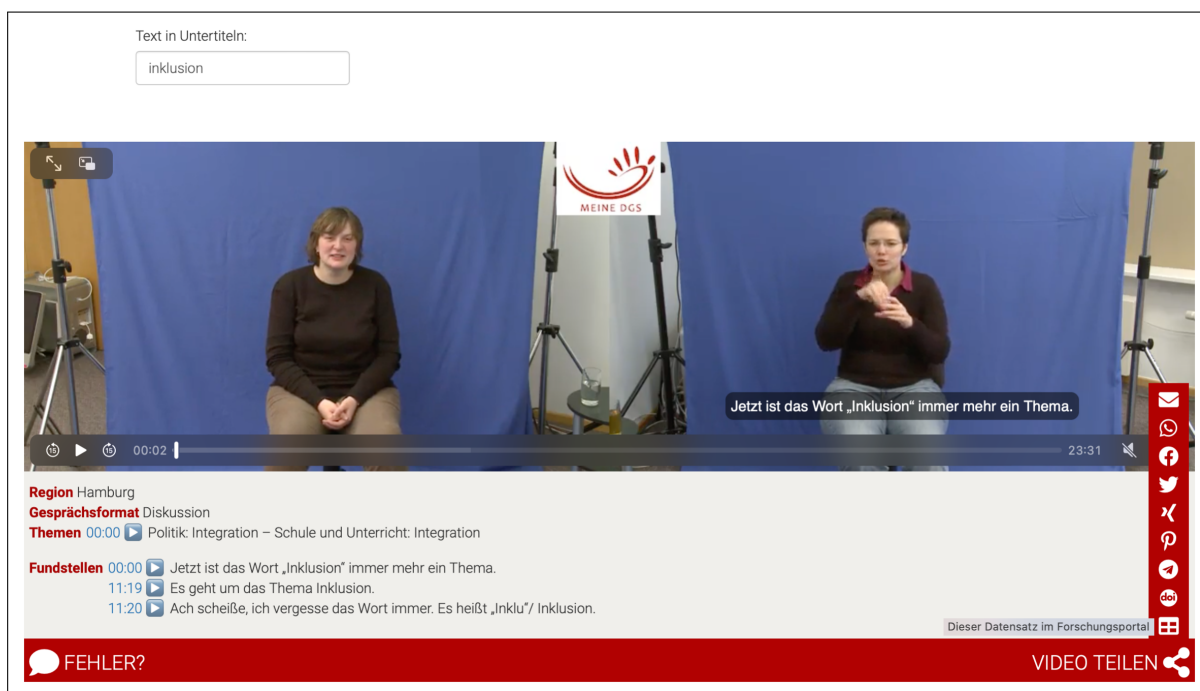


Figure 1: *MY DGS* video with subtitle search field above the video and metadata (region, format, topic) and subtitle search results below. In the lower right corner are buttons for jumping to *MY DGS – annotated* and sharing via social media or email.

translations by entering only the wildcard by itself in the search field.

A new search field was also added to the topics filter (“Alle Themen”) which now lists, in addition to the 33 main topics, more than 570 keywords (previously only included in *MY DGS – annotated*). Browsing keywords and using the search function helps to quickly find videos of interest.

3.3. Changes to *MY DGS – annotated*

MY DGS – annotated receives improvements to its web interface and the amount of information and data it provides.

3.3.1. Type Entries

The pages for individual type entries have received several improvements, particularly regarding information provided for subtypes, i.e. specific meanings of a sign (see “double glossing” in Konrad et al. (2022) for details). An example of the subtype section of a type entry can be seen in Figure 2.

Where a mouthing typically accompanies a sign meaning, this is now specified in the subtype entry. Where available, this is accompanied by a video of the sign’s citation form with the appropriate mouthing, which functions as a supplement to the citation form video without mouthing that is provided for the main type.

Subtypes now also provide translational equivalents for the given meaning of a sign. Depend-

ing on the chosen interface language, these are translations into German or English. Translational equivalents are sourced from the lexical inventory of the DGS Corpus iLex database and are specified based on the complete reference corpus as well as other information sources, so they may include translations that are not based on tokens of the Public DGS Corpus.

The HamNoSys notations provided for sign types have undergone a major quality assurance revision, improving their quality and consistency.

At the bottom of the type entry page, a table of downloadable data for the type and its subtypes is provided. Similar to the download table for transcripts, it provides annotation files in iLex XML, ELAN and SRT formats, video files of the available camera perspectives, and pose information.

3.3.2. Types List

The types list receives a new search feature that lets users filter the list of type glosses by entering (partial) gloss strings or translational equivalents. The search also allows filtering by phonetic attributes via HamNoSys notation.² Gloss and HamNoSys strings can be combined to narrow down the search further, as can be seen in Figure 3.

²To produce HamNoSys symbols, one can use the platform independent HamNoSys editor at <https://www.sign-lang.uni-hamburg.de/hamnosys/input/>

INKLUSIV1 (26 Tokens)

Mund: inklusiv

🎥
frontal
schräg von vorne
seitlich
von oben

Übersetzungsäquivalente: Inklusion; inklusiv; inklusive

▶ 1178939 hh07 | 31-45f Jetzt ist das Wort „Inklusion“ immer mehr ein Thema.

r	MEHR3	AUCH3A*	WORT3	INKLUSIV1
l				
m	mehr mehr	auch		inklusion

Figure 2: Excerpt of a type entry page in *MY DGS – annotated*, showing the subtype INKLUSIV1. Subtypes now specify their typical mouthing (here: *inklusi*v) and provide a mouthing-specific citation form video (when available) and translational equivalents in German or English, depending on the interface language.

In addition to the ordering by type gloss, a new ordering by translational equivalents is provided. As discussed in Section 3.3.1, these are German or English translations for subtypes of a sign. The type list gathers all translations, sorted alphabetically, and for each translation lists all subtypes that specify it as a semantic equivalent.

Subtypes with multiple translational equivalents are listed repeatedly, as can be seen in Figure 4. For instance, INKLUSIV1 is listed for “Inklusion” (inclusivity), “inklusiv” (inclusive) and “inklusive” (including), while INKLUSIV2 and INKLUSIV3 have only the latter two as translational equivalents. This types list ordering can also be filtered, in this case based on (partial) strings of translational equivalents.

3.3.3. Transcripts

The transcript overview has been extended to provide the numeric identifier of each transcript that are used in corpus filenames and DOIs, making it easier for users to find the correct transcript for a downloaded file.

The overview also received a visual indicator for whether lemmatisations, translations or only video material is available for a given transcript.

The transcript viewer has been updated to support the connections to other portals described in Section 4.

3.3.4. Data Collection Formats

The pages describing individual data collection formats now provide a link to the corresponding entry in the *Sign Language Dataset Compendium* (Kopf et al., 2022), a resource compiling information on corpora and lexical resources for sign languages, including a list of commonly used collection tasks and which corpora include them.

3.3.5. Pose information

MY DGS – annotated provides pose information for computational analysis of corpus data. Pose information represents participants as a set of automatically determined keypoints, image coordinates of specific points on the body. Until now, all pose information provided by *MY DGS – annotated* was generated using the OpenPose pose recognition model (Cao et al., 2021; Simon et al., 2017) (see Hanke (2019) and Schulder and Hanke (2020) for details). With release 4, additional outputs generated using MediaPipe (Lugaresi et al., 2019) and Apple Vision Framework³ pose models are made available. As a result, several alternative pose representations are now available for every corpus recording, including representations with 3D estimates for keypoint locations.

³<https://developer.apple.com/documentation/vision/>

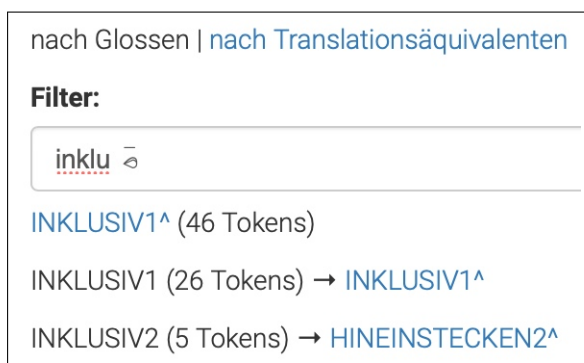


Figure 3: *MY DGS – annotated* gloss search that combines text string search with a phonetic restriction to only show glosses whose sign type uses a hand with fingers that are pinched together while stretched out, signified by HamNoSys notation $\bar{\circ}$.

3.4. Changes to MY DGS – ANNIS

The *MY DGS – ANNIS* portal has been available since 2022, and allows researchers to search the German or English version of the Public DGS Corpus (Isard and Konrad, 2022). Queries can cover multiple transcripts and/or multiple annotation tiers, and can also include corpus metadata. Search results are shown with annotations displayed as a horizontal grid, linked to the video file for a transcript. Export of results in csv format is possible if further processing of the results is required.

Release 4 of the Public DGS Corpus is made available through *MY DGS – ANNIS*, but access to the release 3 data is maintained, as is the case with subsequent Public DGS Corpus releases through other portals. For a demonstration of the *MY DGS – ANNIS* interface, see Figure 6 and Section 5.

With release 4, we add new features to *MY DGS – ANNIS*, including an additional tier of keywords linked to the existing tiers, so it is possible for example to search for vocabulary which occurs particularly often during discussions of a particular topic. Release 4 also introduces links to the other Public DGS Corpus portals as described in Section 4.

MY DGS – ANNIS has proved to be a popular resource among the sign language research community, but the richness of the annotation data combined with the need to learn the basics of the query language AQL can be a barrier for some researchers. We have therefore introduced a Query Wizard which allows users to build up a query using an interface which requires basic knowledge of the annotations of the Public DGS Corpus, but no prior knowledge of AQL. A detailed description of the Query Wizard can be found in Isard (2024). The Query Wizard is compatible with each *MY DGS – ANNIS* dataset (each covering a specific corpus release and annotation language) and will be updated for use with future corpus releases. This has the

Inhalt	INHALT1 (27 Tokens) → HINEINSTECKEN2^ INHALT3 (134 Tokens) → HINEINSTECKEN3^
Inklusion	INKLUSIV1 (26 Tokens) → INKLUSIV1^
inklusive	INKLUSIV1 (26 Tokens) → INKLUSIV1^ INKLUSIV2 (5 Tokens) → HINEINSTECKEN2^ INKLUSIV3 (3 Tokens) → HINEINSTECKEN3^
inklusive	INKLUSIV1 (26 Tokens) → INKLUSIV1^ INKLUSIV2 (5 Tokens) → HINEINSTECKEN2^ INKLUSIV3 (3 Tokens) → HINEINSTECKEN3^
Inkorporation	INKORPORATION1 (4 Tokens) → INTEGRATION1^
Inliner	INLINER1A (5 Tokens) → SCHLITTSCHUHLAUFEN1B^

Figure 4: Excerpt of *MY DGS – annotated* type list, grouped by translational equivalent. Subtypes with multiple possible translations are repeated for each of them.

benefit of allowing researchers to seamlessly transition between versions, even when structural differences between the datasets necessitate changes to the AQL queries.

3.5. Introducing MY DGS – iLex

Users with experience in using the annotation software iLex can now use it to access a read-only version of the Public DGS Corpus through the *MY DGS – iLex* portal. This representation most closely matches the internal annotation environment of the DGS-Korpus project. As such it provides strong support for advanced structures of the corpus, such as the type hierarchy and double token tags (Konrad et al., 2022).

Users familiar with interfacing with PostgreSQL databases can also directly access the iLex database of the *MY DGS – iLex* portal.

4. Connecting the Portals

The Public DGS Corpus is a resource for a variety of groups, such as the deaf community, linguistic researchers and sign language educators and students. To serve these different groups, different corpus portals were optimised for the needs of specific groups and their use cases. Yet as the corpus grew and evolved, it became clear that the interests of the individual groups were not necessarily limited to a single portal. Rather, the optimised experience of one portal could lead users to become familiar enough with its contents to wish to explore additional facets better served by another portal.

For instance, *MY DGS – annotated* was initially intended as purely targeting the international research community and therefore its interface was available only in English. However, feedback from

From	To	Link
<i>MY DGS</i>	<i>MY DGS – annotated</i>	to the same transcript in the other portal
<i>MY DGS – annotated</i>	<i>MY DGS</i>	to the same transcript in the other portal
	<i>MY DGS – ANNIS</i>	from each type/subtype token, translation or mouthing annotation to the search result for the corresponding item
	<i>MY DGS – iLex</i>	from each type/subtype token, translation or mouthing annotation to the corresponding iLex entry
<i>MY DGS – ANNIS</i>	<i>MY DGS – annotated</i>	from search results to corresponding position in transcript or entry in the types list
	<i>MY DGS – iLex</i>	from each item in search results to the corresponding iLex type or token.
<i>MY DGS – iLex</i>	<i>MY DGS</i>	to the same transcript
	<i>MY DGS – annotated</i>	from each type/subtype token, translation or mouthing annotation to the corresponding timecode or entry in the types list
	<i>MY DGS – ANNIS</i>	from each type/subtype token, translation or mouthing annotation to the search result for the corresponding item

Table 1: Table of the links between the four Public DGS Corpus portals

the deaf community⁴ showed that they were interested in also exploring the research data aspect of the Public DGS Corpus, so in release 2 a German interface was added to provide better accessibility (Hanke et al., 2020).

With release 4, the portals turn from separate resources into a network of interfaces to explore the Public DGS Corpus and be combined for new emerging use cases. This is a continuation of our interconnectivity efforts that started with the inclusion of links to DW-DGS and several specialist dictionaries on type entry pages in release 3.

Depending on the portal, different kinds of links to the other portals exist. These connections are described in Table 1. An illustration of the network of connections can also be seen in Figure 5.

Users can move from within a transcript to the same transcript in other portals. Most commonly this leads to the beginning of the transcript, although depending on the structure of the in- and outgoing portals, jumping to a specific timecode or token might also be possible. Connections for types are also available. These lead either to type entries of the other portal or, in the case of *MY DGS – ANNIS* to a query listing all token occurrences of the type.

For some concrete examples of how the interconnected portals can be used, see the use cases described in Section 5.

⁴Such feedback was received via the project’s dedicated “focus group”, a team of deaf experts providing advisory and outreach capacities that connect the project to communities of different regions (Prillwitz et al., 2008).

5. Use Cases

In this section we describe some use case scenarios to illustrate ways to start a search in the Public DGS Corpus and how to profit from the connections between portals.

5.1. From topics and keywords to meanings and tokens

Let us assume that you are interested in whether inclusivity was already a topic of debate in the early 2010s. You can search for the keyword “Inklusion” in the topics of *MY DGS* (filter “Alle Themen”) where you will find six videos. Opening a video and using the search field “Text in Untertiteln” (text in subtitles) you will find the German translations containing “Inklusion”. Using the timecode link you can direct the video to the starting point of the translation.

Not all occurrences of inclusivity occur in transcripts which have the topic “Inklusion”, so searching for “inklu” in the subtitles search field will return a larger number of hits, namely twelve videos. The subtitles contain 16 translations with “Inklusion” or the adjective “inklusive” or the verb “inkludieren”, all in the sense of ‘concerning inclusivity’, and 10 translations with the adjective “inklusive” in the English sense of “all inclusive”.

Moving on to *MY DGS – annotated*, you can search the types by translational equivalents⁵ in order to find subtypes which are associated with German words beginning with “inklu”. For example,

⁵https://meine-dgs.de/ling/meanings_de.html

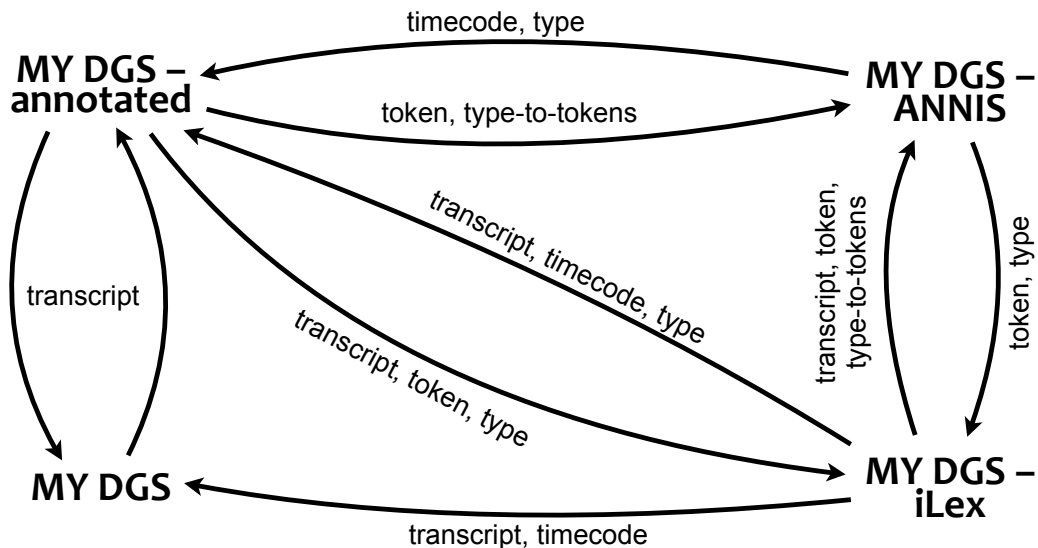


Figure 5: Diagram of the connections between the four Public DGS Corpus portals. *Transcript* denotes links to a given video transcript, *timecode* to a specific time point in the video, *token* to a specific token or mouthing/translation annotation in the transcript, *type* to an entry in the types list, *type-to-token* to a search of all tokens that are instances of the given type.

INKLUSIV⁶ has the equivalents “Inklusion; inklusiv; inklusive” (inclusivity; inclusive; included). For each type you can then inspect the translations manually to find the German translations containing “Inklusion” or the adjective “inklusiv” in the sense of ‘concerning inclusivity’, but not in the sense of ‘all inclusive’. From the type entry you can also use the links to *MY DGS – ANNIS* and *MY DGS – iLex* to continue and refine your search.

In *MY DGS – ANNIS*, you can also use AQL query expressions to look for links between different annotation tiers. Instead of looking at the type entry, you could, for example, use [Query 1](#) to search for translations with “inkl” linked to glosses which have a Mouthing which also starts with “inkl”. This returns 19 matches, as shown in [Figure 6](#).

- (1) Gloss ->ident Mundbild=/inkl.*/ & Deutsch=/.*[il]nklu.*/ & #1 ->ident #3

From the matches in *MY DGS – ANNIS* you can jump to either the type entries or the relevant transcript sections of *MY DGS – annotated* and *MY DGS – iLex*, as described in [Section 4](#).

In *MY DGS – iLex* you can use a customized SQL query searching the German translations for “inkl” but excluding “alles inklusiv” and “inklusive alle”. You can also make use of a number of predefined SQL functions of *MY DGS – iLex*, such as the `tag_to_glossstring()` function, which takes a translation tag ID and outputs the sequence of token glosses that are covered by the translation. In our search, this function could be added to the

⁶<https://doi.org/10.25592/dgs.corpus-4.0-type-51514#type51515>

SQL query to not only see the relevant translations, but also the textual representation of the underlying signed utterance to make a preliminary confirmation regarding the relevance of the sentence.

5.2. Idiomatic phrases

Looking for collocations is one step in the lexical description of words and signs. This may lead to the discovery of multi-word expressions or even idiomatic phrases. The latter seem to be very rare in sign languages ([Johnston and Ferrara, 2012](#); [Wilkinson et al., 2023](#)). Some examples of phrases in DGS can be found in the Digital Dictionary of DGS (DW-DGS)⁷. For example, entry 262⁸ identifies the phrase WARM1A⁹ GROUP1A¹⁰ which is described as meaning *cordial* or *communal* in the sense of cohesion and interaction in groups of people. The DW-DGS entry provides three examples of the phrase being signed in the DGS Corpus.

Based on the information in DW-DGS, one could assume that “WARM1A⁹ GROUP1A¹⁰” is a fixed phrase with a semantically idiomatic status (cf. [Wilkinson et al., 2023](#)). To confirm this with the Public DGS Corpus data, one can run searches in *MY DGS – ANNIS*. The strength of this corpus search tool is that you can easily define the context of a token,

⁷See [Langer et al. \(2022\)](#) for a detailed description of information types in the DW-DGS.

⁸<https://www.sign-lang.uni-hamburg.de/korpusdict/bags/bag262.html>

⁹<https://doi.org/10.25592/dgs.corpus-4.0-type-13170>

¹⁰<https://doi.org/10.25592/dgs.corpus-4.0-type-13141>

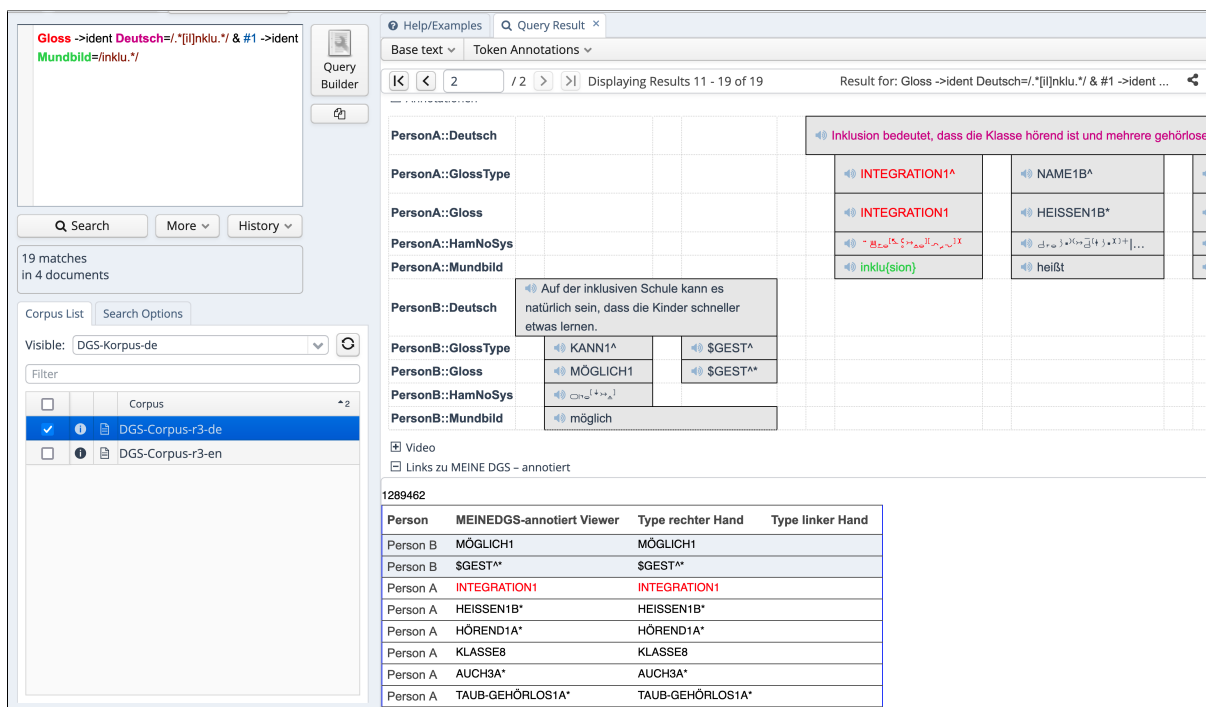


Figure 6: MY DGS – ANNIS search for German translation containing “[li]nklu” and Mundbild including “inklu”.

i.e. the preceding or following neighbours, specifying the distance between the search item and its neighbours. The numbers and letters after the gloss name are not specified in this and the following queries in order to retrieve all combinations of variants of types with the gloss names WARM and GROUP. For details on how you can build these AQL queries with the ANNIS Query Wizard see [Isard \(2024\)](#).

First, you can check the the sequence WARM GROUP using [Query 2](#), for which there are 8 matches in the Public DGS Corpus.

- (2) GlossType=/WARM.*/.GlossType
GlossType=/GROUP.*/
& Gloss ->ident English & #3 ->ident #1

Second, you can check the reverse sign order, GROUP WARM, using [Query 3](#), for which there is 1 match.

- (3) GlossType=/GROUP.*/.GlossType
GlossType=/WARM.*/
& Gloss ->ident English & #3 ->ident #1

Third, you can search for co-occurrences of WARM and GROUP in either order and with one to four tokens in between using [Query 4](#), for which there are 10 matches.

- (4) GlossType=/GROUP.*/^GlossType,2,5
GlossType=/WARM.*/
& Gloss ->ident English & #3 ->ident #1

A closer look at the data has shown us that WARM GROUP is not just a fixed combination, but may also occur in reverse order and interrupted by other lexical elements.

6. Conclusion

We presented release 4 of the Public DGS Corpus. It introduces one more hour of recordings from the DGS Corpus, new pose data and a new portal, *MY DGS – iLex*. All portals receive improvements such as added information, new search capabilities and other interface refinements.

A major change in release 4 is the added interconnectivity between portals. Each portal offers a number of ways to jump to other portals, allowing for a more dynamic use of the resources, combining each of their strengths.

7. 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.

8. Bibliographical References

- Zhe Cao, Ginés Hidalgo Martinez, Tomas Simon, Shih-En Wei, and Yaser A. Sheikh. 2021. [OpenPose: Realtime multi-person 2D pose estimation using part affinity fields](#). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):172–186.
- Wallace L. Chafe, editor. 1980. *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production*. Number 3 in *Advances in discourse processes*. Ablex Publishing Corporation, Norwood, New Jersey, USA.
- Thomas Hanke. 2002. [iLex - a tool for sign language lexicography and corpus analysis](#). In *3rd International Conference on Language Resources and Evaluation (LREC 2002)*, pages 923–926, Las Palmas, Canary Islands, Spain. European Language Resources Association (ELRA).
- Thomas Hanke. 2019. [Processing DGS-Korpus data with OpenPose on the Hamburg High Performance Cluster](#). Project Note AP04-2018-01, DGS-Korpus project, IDGS, Hamburg University, Hamburg, Germany.
- Thomas Hanke, Marc Schulder, Reiner Konrad, and Elena Jahn. 2020. [Extending the Public DGS Corpus in size and depth](#). 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 75–82, Marseille, France. European Language Resources Association (ELRA).
- Amy Isard. 2024. [Building your query step by step: A Query Wizard for the MY DGS – ANNIS portal of the DGS Corpus](#). In *Proceedings of the LREC2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources*, Turin, Italy. European Language Resources Association (ELRA).
- Amy Isard and Reiner Konrad. 2022. [MY DGS – ANNIS: ANNIS and the Public DGS Corpus](#). In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 73–79, Marseille, France. European Language Resources Association (ELRA).
- Elena Jahn, Reiner Konrad, Gabriele Langer, Sven Wagner, and Thomas Hanke. 2018. [Publishing DGS Corpus data: Different formats for different needs](#). In *Proceedings of the LREC2018 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community*, pages 83–90, Miyazaki, Japan. European Language Resources Association (ELRA).
- Trevor Johnston and Lindsay Ferrara. 2012. [Lexicalization in signed languages: When is an idiom not an idiom?](#) In *Selected Papers from UK-CLA Meetings*, volume 1, pages 229–248. UK-CLA.
- Reiner Konrad, Thomas Hanke, Gabriele Langer, Susanne König, Lutz König, Rie Nishio, and Anja Regen. 2022. [Öffentliches DGS-Korpus: Annotationskonventionen / Public DGS Corpus: Annotation conventions](#). Project Note AP03-2018-01, DGS-Korpus project, IDGS, Hamburg University, Hamburg, Germany.
- Maria Kopf, Marc Schulder, and Thomas Hanke. 2022. [The Sign Language Dataset Compendium: Creating an overview of digital linguistic resources](#). In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 102–109, Marseille, France. European Language Resources Association (ELRA).
- Thomas Krause and Amir Zeldes. 2016. [ANNIS3: A new architecture for generic corpus query and visualization](#). *Digital Scholarship in the Humanities*, 31(1):118–139.
- Gabriele Langer, Anke Müller, Felicitas Otte, and Sabrina Wähl. 2022. [Information types and use cases](#). Project Note AP10-2021-02, DGS-Korpus project, IDGS, Hamburg University, Hamburg, Germany.
- Gabriele Langer, Anke Müller, Sabrina Wähl, Felicitas Otte, Lea Sepke, and Thomas Hanke. 2024. [Introducing the DW-DGS – the digital dictionary of DGS](#). In *Proceedings of the LREC2024 11th Workshop on the Representation and Processing of Sign Languages: Evaluation of Sign Language Resources*, Turin, Italy. European Language Resources Association (ELRA).
- Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. 2019. [MediaPipe: A framework for perceiving and processing reality](#). In *Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR) 2019*.
- Siegmund Prillwitz, Thomas Hanke, Susanne König, Reiner Konrad, Gabriele Langer, and

- Arvid Schwarz. 2008. [DGS Corpus project – development of a corpus based electronic dictionary German Sign Language / German](#). In *Proceedings of the LREC2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, pages 159–164, Marrakech, Morocco. European Language Resources Association (ELRA).
- Marc Schulder and Thomas Hanke. 2020. [Open-Pose in the Public DGS Corpus](#). Project Note AP06-2019-01, DGS-Korpus project, IDGS, Hamburg University, Hamburg, Germany.
- Marc Schulder and Thomas Hanke. 2022. [How to be FAIR when you CARE: The DGS Corpus as a case study of open science resources for minority languages](#). In *13th International Conference on Language Resources and Evaluation (LREC 2022)*, pages 164–173, Marseille, France. European Language Resources Association (ELRA).
- Sehen statt Hören. 2006. Episode 1291: Die Haushaltshilfe. Broadcast by Bayrischer Rundfunk.
- Tomas Simon, Hanbyul Joo, Iain Matthews, and Yaser Sheikh. 2017. [Hand keypoint detection in single images using multiview bootstrapping](#). In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4645–4653, Honolulu, Hawaii, USA.
- Erin Wilkinson, Ryan Lopic, and Lynn Hou. 2023. [Usage-based grammar: Multi-word expressions in American Sign Language](#). In Terry Janzen and Barbara Shaffer, editors, *Signed Language and Gesture Research in Cognitive Linguistics*, pages 357–388. De Gruyter Mouton, Berlin, Boston.